Digital Filter Design by Approximate Time-Domain Modeling

by

Ali Saud Abbas Al-Ahmadi

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

June, 1994

DIGITAL FILTER DESIGN BY APPROXIMATE TIME-DOMAIN MODELING

BY

ALI SAUD ABBAS AL-AHMADI

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

JUNE 1994

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS
DHAHRAN 31261, SAUDI ARABIA

COLLEGE OF GRADUATE STUDIES

This thesis, written by ***ALI SAUD ABBAS AL-AHMADI*** under the direction of his Thesis Advisor and approved by his Thesis Committee, has been presented to and accepted by the Dean of the College of Graduate Studies, in partial fulfillment of the requirements for the degree of ***MASTER OF SCIENCE in ELECTRICAL ENGINEERING.***

THESIS COMMITTEE

Dr. M. Bettayeb (Advisor)

Dr. J. Bakhashwain (Member)

Dr. S. Penbeci (Member)

Dr. S. Al-Baiyat (Member)

Department Chairman

Dean, College of Graduate Studies

ACKNOWLEDGMENT

First of all, praise be to "ALLAH" whose help made it possible to complete this work. Acknowledgment is due to King Fahd University of Petroleum and Minerals for support of this research.

I wish to thank my thesis advisor Dr. Maamar Bettayeb for his patience and careful guidance during the course of work. I am very grateful to my other committee members Dr. Jamil Bakhashawain, Dr. Suleyman Penbeci, and Dr. Samir Al-Baiyat for their helpful remarks and cooperation.

I would like to thank my brothers Yasser Al-Ghamdi and Suleyman Al-Mani for their great help in PC work.

I am greatly indebted to all my family: father, mother, brothers, sisters, wife, and my little sons for their love, inspiration, and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ملخص الرسالة :

اسم الطالب : علي سعود الأحمدي

عنوان الرسالة : تصميم المرشحات الرقمية بواسطة التمثيلات الزمنية التقريبية

التخصص : هندسة كهربائية

تاريخ الشهادة : محرم ١٤١٥هـ الموافق يونيو ١٩٩٤م

هذه الرسالة تتكون من قسمين رئيسيين :

القسم الأول : يتناول تطبيق طريقتين من الطرق المثلى بمقياس هانكل لتصميم وتقريب المرشحات الرقمية ذات الإستجابة النبضية اللامنتهية .وهاتان الطريقتان هما الطريقة المثلى بمقياس هانكل بدون الحد الثابت والطريقة الأخرى هي الطريقة المثلى بقياس هانكل مع الحد الثابت . وقد تم مقارنة هاتين الطريقتين بعدة طرق أخرى لتصميم هذا النوع من المرشحات ، وهذه الطرق هي : طريقة تقليل مربع الخطأ ، وطريقة كاراثودري فيجير ، وطرق التقريب المتوازنة ، وكذلك طريقة كينج ، وطريقة كيمورا والطريقتين الأخيرتين تعد طرق قريبة من المثالية بمعيار هانكل وقد تم استخلاص نتائج مهمه من هذه الدراسة المقارنة في نهاية البحث .

القسم الثاني : يتناول تطبيق الطريقة الثنائية للتقريب على الطرق المذكورة في القسم الأول ، وقد استنبطت معادلات عامة لتطبيق الطريقة الثنائية للتقريب على مجال واسع من الطرق المتعلقة بتصميم المرشحات الرقمية في المجال الزمني ، وقد أظهرت هذه الطريقة كفاءة عالية جداً لتقريب القيمة المطلقة للمرشحات الرقمية في المجال الترددي باستخدام مرشحات رقمية ذات قوى أسية بسيطة بالمقارنة مع الطريقة الأحادية للتقريب .

درجة الماجستير في العلوم الهندسية

جامعة الملك فهد للبترول والمعادن

الظهران - المملكة العربية السعودية

محرم ١٤١٥ هـ الموافق يونيو ١٩٩٤م

ABSTRACT

Title : **DIGITAL FILTER DESIGN BY APPROXIMATE TIME-DOMAIN MODELING.**

By : **Ali Saud Al-Ahmadi.**

Major Field : **Electrical Engineering.**

Date : **June, 1994.**

This thesis consists of two main parts. In the first part, two optimal Hankel methods: optimal Hankel without a D-term developed by Bettayeb and optimal Hankel with a D-term developed by Glover, are applied to IIR (infinite impulse response) digital filter design and approximation. The results obtained from optimal Hankel methods are compared to the results obtained from other time-domain design methods: Least squares methods, CF method, balanced approximation methods, and suboptimal Hankel methods. Important conclusions are derived from this comparative study. Moreover, a slight modification is made to one of the optimal Hankel methods to improve its efficiency. The D-term which is neglected in this method is forced to be equal to the dc term of the impulse response. This improvement is shown by examples.

In the second part, the two-sided approximation technique for IIR digital filter design is applied to the above methods. General formulae are derived so that any time-domain IIR digital filter design method could be applied using the two-sided approximation. It is shown through different examples that this technique gives a very efficient approximations to the desired magnitude response with low-order IIR digital filters compared to the one-sided approximation technique.

CHAPTER 1

INTRODUCTION

1.1 General

Digital filters play an important role in all digital signal processing applications such as data communication, sonar and radar processing, speech processing and seismic signal processing. Digital filters are of two main types: finite impulse response (FIR) filter and infinite impulse response IIR filter. Most effective or optimal techniques of IIR digital filter design are based on frequency domain specifications [1]. On the other hand, existing optimal techniques based on time domain are iterative [2] and some of them use nonlinear programming to reach to the desired solution [3,4]. This leads to heavy computation. Moreover, convergence to the optimal design is not always guaranteed [5]. Thus, more efforts should be devoted to obtain simple, noniterative, and optimal or near optimal methods for IIR digital filter design in time domain. A promising approach to this aim is to utilize some well-known techniques in model reduction theory developed in the last decade such as optimal Hankel and balanced approximation. As a matter of fact, there exist many IIR filter design methods based on the above techniques. However, these methods need to be evaluated and compared to other existing methods in the literature.

The FIR filter has many desirable features. It is an all-zero system which is always stable and a more important feature is that FIR filter can be easily designed to obtain linear phase characteristics (LPHC). This feature is of special importance for many applications where frequency dispersion due to nonlinear phase is harmful, e.g. speech processing and data transmission [1]. On the other hand, FIR filters need to be of a high order so that it can meet magnitude response specifications especially with sharp cutoff frequency. This means more complexity and cost. IIR filters are more efficient in approximating magnitude responses with lower order. Thus it is logical to try to incorporate these two features : linear phase property of FIR filters and the efficiency of IIR filters in approximating the magnitude response with low order, in digital filter design.

One possible approach to achieve the above aim is to approximate FIR filter by an IIR filter. Unfortunately, it is impossible to fully satisfy LPHC using this approach since IIR filters can't attain LPHC. Nevertheless, it can achieve approximate LPHC when certain methods are used as will be shown later. Most of time domain methods deal with causal impulse response to design a causal and stable filter. To obtain a causal impulse response of an ideal filter, the Fourier coefficients of the frequency response Fourier expansion are first truncated and then a constant delay is introduced. The effect of this transformation is that the lower order terms of the causal impulse response no longer represent the low-frequency terms of the original Fourier expansion. This constraint was removed by the two-sided approximation technique [6] where the Fourier coefficients of the magnitude response are not shifted and

the approximation is performed on the noncausal truncated Fourier coefficients from which a stable and causal IIR filter is obtained. This technique needs to be investigated more since it was applied using Pade approximation only while other techniques could be also applied. Moreover, the two-sided approximation method was originally designed to deal with the coefficients of the Fourier expansion of the amplitude response to obtain the desired characteristics. Another alternative is to use the impulse response of the desired or approximated digital filter. It is expected that excellent results would be obtained since some of the methods which will be applied instead of Pade approximation are optimal.

1.2 Literature Review

This literature review is rather long since we are dealing with different methods. It is more appropriate if it is divided into subsections.

1.2.1 Pade Approximation

Pade approximation was first discussed by Forbenius [7] over a hundred years ago. Formally, the first paper which introduced Pade approximation to IIR digital filter design was written by Brophy and Salazar [8]. They considered the stability of the Pade approximants and derived sufficient conditions for that. Burrus and parks in [9], although not explicitly stated,

applied Pade approximation in a matrix form. Many extensions and algorithms based on Pade approximation were developed in [10,11].

1.2.2 Least Squares Methods

The least squares error criterion is widely used in different digital signal processing techniques. However, direct minimization requires the solution of a set of nonlinear equations. Several papers have suggested using nonlinear programming methods such as the least squares Taylor method [3] and the Gauss-Newton method [4]. Steiglitz and McBride in [12] proposed an iterative method which weighs and solves a set of overdetermined linear equations in terms of both the numerator and denominator coefficients of the designed filter. Other iterative methods can be found in [2] and [5].

Shanks [13] and Burrus et al. [9] have decoupled the solution for the numerator and denominator coefficients where a two-step approximation is applied. The denominator coefficients, are calculated using least squares approximation and then another least squares approximation is performed to find the numerator coefficients. Prony's method [14] is also a two-step approximation but only the denominator coefficients are obtained using least squares approximation. For the numerator coefficients, they are calculated using a direct recursive formula.

1.2.3 The CF (Caratheodory-Fejer) Method

This method was developed by Gutknecht et. al. in [15]. It is based on an extension due primarily to Takagi [16] of a classical theorem in complex

analysis of Caratheodory and Fejer [17]. An earlier version of this method could be found in [18]. A strong mathematical background of this method is given in [19] which discusses the rational Chebyshev approximation on the unit disk. Another useful paper [20] was written by Gutknecht who applied the CF theorem to obtain rational approximation on a disk, a circle, and an interval. Real rational approximation using CF method could be found in [21].

1.2.4 Balanced Model Reduction

Balanced system realization was a breakthrough in model reduction theory. It was developed independently by Moore [22] and Mullis and Roberts [23]. Moore showed that there exists a nonsingular transformation which transfer the system to a balanced state where the observability grammian and controllability grammian are equal and diagonal. Mullis and Roberts obtained a similar representation but from a different point of view where they considered the minimum round off noise for fixed point digital filters. Pernobo and Silverman [24] studied the stability, controllability, and observability of the balanced model reduction. Davidson [25] proposed another criterion for the selection of states in terms of their input-output contribution. A similar work was done by Kabamba in [26]. Kung in [27] considered the discrete time case. He obtained an equivalent model to Moore's via the Hankel approach. Different versions of Kung's algorithm could be found in [28], [29], and [30]. These methods are suboptimal Hankel methods. Bettayeb et. al. [31,32] utilized the Adamjan-Arov-Krein approximation theory and developed an optimal Hankel norm approximation for finite dimensional systems. Glover

studied the optimal Hankel approximations extensively in [33] where he characterized all optimal Hankel solutions. Also, he developed an algorithm to find the D-term which lowers the error frequency bound to half of the original one (without a D-term).

Fernando and Nicholson [34] proposed a balanced reduced model which is not the case for the above mentioned algorithms where the reduced model is not balanced. Several papers [23,35,36,37] have shown that a balanced reduced model can attain minimum sensitivity to parameter variations. Moreover, it can simultaneously achieve minimum sensitivity to both parameter variation and round off noise if certain conditions are imposed. Al-saggaf in [38] developed a new reduced model which is also balanced but it has an additional features over Fernando and Nicholson model.

1.2.5 Approximation of FIR Filter by IIR Filter

In frequency domain, Heckelmann et.al. [39] proposed a method to approximate the frequency response of a FIR digital filter by an IIR digital filter based on the fact that the frequency response of any stable and causal digital system is completely determined by its real part.

In time domain, more is found about this subject. Based on the R and S array algorithm in model identification, Bednar and Coberly [40] treated the problem of approximating FIR filters by IIR filters. Bednar in [41] proposed a procedure to determine the best order of the IIR filter by estimating the quality of the approximation prior to the design. Balanced approximation was

applied in [42] to approximate FIR filters with LPHC. It was also applied by Beliczynski et al. [43] to approximate FIR by IIR digital filters.

Some methods which were developed to design IIR filters with linear phase are based actually on approximating a LPHC FIR filter by an IIR filter. Such a method is found in [44] where impulse response grammians diagonalization is used for model reduction.

1.2.6 Two-Sided Approximation (Noncausal Approximation)

Chen et. al. [6] were the first to propose a causal and stable approximation for noncausal recursive digital filters. This method was reformulated and extended to two dimensional digital filter design in [45]. Realization of zero phase noncausal filters was first considered by Kormylo et. al. [46] based on two-pass technique. This work was extended by Czarnach [47] for processing infinite length signal processing. Chan et. al. [48] proposed sample-by-sample approach to develop real time realization for noncausal filters. Multiple criterion optimization was applied for the design of noncausal filters with general phase characteristics in [49]. Chan in [50] considered noncausal digital filters design with antisymmetric impulse response.

1.3 Thesis Work

This thesis consists of two main parts. In the first part, an overall comparative study is offered for seven important time domain methods namely:

Pade approximation, least squares method, CF method, internal balanced method, minimum sensitivity method (MS method), suboptimal Hankel methods, and optimal Hankel methods. Many remarks regarding the optimality, stability, and efficiency of the previous methods are included within the thesis. Moreover, a slight modification is suggested to improve the efficiency of one of the optimal Hankel methods. The most important subject of this part is the comparative study performed, using different examples, of these methods in their efficiency of designing IIR digital filters and approximating IIR digital filters and FIR digital filters by lower order IIR digital filters. Magnitude response, phase response, impulse response, L_∞ norm, and least squares error (LSE) are all considered for such approximations. Concrete conclusions are stated at the end of the thesis.

In the second part, the two-sided rational approximation technique for noncausal IIR filter design is considered. This method is applied directly to truncated or finite symmetric or antisymmetric impulse response instead of the amplitude response. Moreover, the original formula of this technique is slightly modified to handle antisymmetric impulse response. It is shown that the error bound given in [6] still holds for antisymmetric impulse response. General explicit formulae are derived for both cases: symmetric and antisymmetric impulse response, which enable the designer to apply any time domain method suitable to be applied for two-sided approximation. A simple algorithm is proposed for that. All of the methods considered in the first part of the thesis are applied and tested using the two-sided approximation technique where excellent results are obtained for magnitude response.

Finally, the above algorithms are implemented in MATLAB and provided as M-files in a user interface form.

1.4 Thesis Orignization

This thesis is organized in six chapters. Chapter 1 is an introductory chapter. Chapter 2 discusses Pade approximation, least squares method, and CF method. Internal balanced realization, MS method, suboptimal and optimal Hankel methods are all introduced in chapter 3. In chapter 4, the two-sided technique is discussed in detail. Examples for illustration and comparison between the above mentioned time domain methods are provided in chapter 5. Conclusions and recommendations for further work are given in chapter 6. Finally, the different routines developed in the MATLAB environment for the various digital filter design techniques treated in this thesis are listed in appendix A.

CHAPTER II

NON-ITERATIVE TECHNIQUES FOR APPROXIMATING IIR DIGITAL FILTERS IN TIME DOMAIN

In this chapter, some of the methods that can be employed to approximate and design IIR digital filters in time domain are discussed. An overview about the theory behind these methods and a description of the algorithms based on it will be provided. Moreover, a brief discussion about advantages and disadvantages of each technique is given at the end of each section.

2.1 Pade Approximation Method

Pade approximation is a simple method which equates h(n) to the first M+N samples of $h_d(n)$ where $h_d(n)$and h(n) stand for the impulse responses of the desired filter and designed filter respectively:

$$H_d(z) = \sum_{n=0}^{\infty} h_d(n) z^{-n} \tag{2.1}$$

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 + \sum_{k=1}^{N} a_k z^{-}} = \sum_{n=0}^{\infty} h(n) z^{-n} \tag{2.2}$$

M and N represent the order of the numerator and denominator of the designed IIR filter.

Practically, the design starts by truncating $h_d(n)$ to obtain a finite set of data of length, say, L.

The optimization criterion used in such techniques is the least squares criterion [14] :

$$\varepsilon = \sum_{0}^{L-1} [h_d(n) - h(n)]^2 \tag{2.3}$$

The error ε or least squares error (LSE) is minimized with respect to the coefficients a_k and b_k. However, h(n) is generally a non-linear function of the filter parameters and consequently this will lead to a set of non-linear equations which have to be solved to minimize ε [28]. This could be seen from the following relation:

$$\sum_{k=0}^{N} a_k h_{i-k} = \begin{cases} b_i & i = 0, 1, .., M \\ 0 & i > M \end{cases} \tag{2.4}$$

If M+N is set to be equal to L-1, then, eq.(2.4) above could be solved for a_k and b_k with h(n) perfectly matched to $h_d(n)$ for $0 \le n \le M+N$ [9]. This is explained through the difference equation of H(z) as follows:

$$H(z) = \frac{Y(z)}{X(z)}$$

$$\begin{aligned} y(n) = &- a_1 y(n-1) - a_2 y(n-2) - .....- a_N y(n-N) + b_0 x(n) \\ &+ b_1 x(n-2) +....+ b_M x(n-M) . \end{aligned} \tag{2.5}$$

If $x(n) = \delta(n)$, then

$$y(n) = h(n) \tag{2.6}$$

and eq.(2.5) becomes:

$$\begin{aligned} h(n) = &- a_1 h(n-1) - a_2 h(n-2) - \ldots - a_N h(n-N) + b_0\delta(n) \\ &+ b_1 \delta(n-1) + \delta(n-M)\, b_M (n-M) . \end{aligned} \tag{2.7}$$

Since $\delta(n-k) = 0$ except for $n = k$, eq.(2.7) is written as follows:

$$h(n) = - a_1 h(n-1) - a_2 h(n-2) - \ldots - a_N h(n-N) + b_n \qquad 0 \le n \le M \tag{2.8}$$

and for $n > M$:

$$h(n) = - a_1 h(n-1) - a_2 h(n-2) - \ldots - a_N h(n-N) \tag{2.9}$$

If h(n) is set equal to $h_d(n)$ for the first M+N samples, then eq.(2.9) could be solved to obtain a_k. Then, b_k is found by substituting a_k in eq.(2.8). A reformulating of the above solution in matrix notation can be found in [9].

Although Pade approximation method is simple and non-iterative, it has two main disadvantages. Firstly, the stability of the designed filter is not guaranteed unless a certain condition is satisfied [8]. Secondly, the higher the order of the desired filter, the more complex is the designed filter. In other words, to have a good approximation using Pade technique, M+N should be equated to L-1. Thus, if L is high, then M and N will be also high. Actually, this is what makes Pade technique impractical for digital filter design. Moreover, if $M+N < L-1$, then the first M+N samples of h(n) will be perfectly matched to $h_d(n)$. However, h(n) with $n > M+N$ will have mostly a serious error which means that the method is not optimal (actually, far from optimality) in such cases .

2.2 Least Squares Methods

The least-squares methods presented here are time-domain design techniques of IIR digital filters. One starts from a finite segment of the impulse response to design an IIR filter whose Fourier coefficients approximate optimally in least-squares sense the desired impulse response over a finite range. By approximating a finite segment of the impulse response, it is hoped that the truncated samples would be approximated within an acceptable range of error.

2.2.1 Problem Formulation

Given $h_d(n)$, find a_k and b_k of H(z) :

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1+\sum_{k=1}^{N} a_k z^{-k}} = \sum_{n=0}^{\infty} h(n) z^{-n} \qquad (2.10)$$

which best minimize the least-squares error criterion given in (2.1).

As mentioned previously, solving for a_k and b_k simultaneously is a difficult task. However, this can be avoided by reformulating the problem as a two step approximation problem developed by Shank in [13]. The first step finds a_k and then, the second step uses a_k to calculate b_k. In the following two sections, two least-squares methods are presented: Prony's method and Shank's method.

2.2.2 Prony Method

Eq.(2.9) can be rewritten as follows:

$$h(n)=\sum_{k=1}^{N} a_k h(n-k) \qquad n > M \qquad (2.11)$$

Eq.(2.11) could be used to find a_k. However, h(n) is not known beforehand. A possible solution is to replace h(n) by $h_d(n)$ to get an approximate solution as follows [28]:

$$h_d(n) \approx \sum_{k=1}^{N} a_k h_d(n-k) \qquad n > M \qquad (2.12)$$

or in matrix notation:

$$\mathbf{h}_d \approx \mathbf{H}_d \mathbf{a} \qquad (2.13)$$

where:

$$\mathbf{h}_d = \begin{bmatrix} h_d(M+1) \\ h_d(M+2) \\ \cdot \\ \cdot \\ h_d(L-1) \end{bmatrix}, \quad \mathbf{H}_d = \begin{bmatrix} h_d(M) & h_d(M-1) & \cdots & h_d(M+1-N) \\ h_d(M+1) & h_d(M) & \cdots & h_d(M+2-N) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ h_d(L-2) & h_d(L-3) & \cdots & h_d(L-1-N) \end{bmatrix}$$

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_N \end{bmatrix} .$$

Eq.(2.13) is mostly an overdetermined system and therefore there is no exact solution for **a** However, a least-squares solution exists for eq.(2.11). It

requires the minimization of the following sum of squares of the approximation error:

$$\sum_{n=M+1}^{L-1} [h_d(n) - \sum_{k=1}^{N} a_k h_d(n-k)]^2 \qquad (2.14)$$

over all possible choices of a_k. It can be shown that the optimal choice of a_k is obtained by solving the following linear system of normal equations: [28]

$$\mathbf{H^T H a = H^T h} \qquad (2.15)$$

or:

$$\begin{bmatrix} \phi(1,1) & \phi(1,2) & \dots & \phi(N,1) \\ \phi(2,1) & \phi(2,2) & \dots & \phi(N,2) \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ \phi(N,1) & \phi(N,2) & \dots & \phi(N,N) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_N \end{bmatrix} = \begin{bmatrix} \phi(1,0) \\ \phi(2,0) \\ . \\ . \\ . \\ \phi(N,0) \end{bmatrix} \qquad (2.16)$$

where :

$$\phi(i,j) = \sum_{n=M+1}^{L-1} h_d(n-j) h_d(n-i) \qquad (2.17)$$

The next step is to find b_k using eq.(2.8) which can be written as:

$$b_n = h_d(n) - \sum_{k=1}^{n} a_k h_d(n-k) \quad 0 \le n \le M \ . \qquad (2.18)$$

Although Prony's method is practical and more effective than the previous method, it is not optimal. Actually, it gives good estimates for the a-coefficients but this not the case for the b-coefficients [14]. This disadvantage is handled in the next section.

2.2.3 Shank Method

Shank [13] and Burrus et al. [9] solved the above problem by proposing another least-squares approximation to get b_k. To begin with, this method finds a_k as described in the last section. Then, it finds b_k by looking at H(z) in the following way:

$$H(z)=B(z)A^{-1}(z) \tag{2.19}$$

This equation indicates that the impulse response of the IIR filter can be thought of as the output of the all-pole system 1/A(z) driven by the finite sequence of the b-coefficients [28]. Thus :

$$h(n)=\sum_{k=0}^{M} b_k g(n-k) \qquad n \geq 0 \tag{2.20}$$

where g(n) is the impulse response of $G(z)= A^{-1}(z)$. Then, taking the inverse z-transform of the relation $A(z)G(z)=1$, the following relation is obtained:

$$g(n)-\sum_{k=1}^{N} a_k g(n-k) = \delta(n) \tag{2.21}$$

From the above relation, g(n) can be calculated as follows:

$$g(0)=1$$
$$g(n)=\sum_{k=1}^{N} a_k g(n-k) \qquad \text{for } 0 < n < L \tag{2.22}$$

After getting g(n) , b_k is found such that:

$$h_d(n) \approx \sum_{k=0}^{M} b_k g(n-k) \qquad \text{for } 0 \le n < L \tag{2.23}$$

The rest follows the same argument when getting a_k. The minimal coefficient b_k can be found by solving the following system of equations:

$$\mathbf{G^T G b = G^T h} \tag{2.24}$$

where:

$$\mathbf{G} = \begin{bmatrix} g(0) & 0 & . & . & 0 \\ g(1) & g(0) & . & . & 0 \\ . & . & . & . & . \\ g(M) & g(M-1) & . & . & g(0) \\ . & . & . & . & . \\ g(L-1) & g(L-2) & . & . & g(L-M-1) \end{bmatrix}$$

$$\mathbf{h} = \begin{bmatrix} h_d(0) \\ h_d(1) \\ . \\ . \\ h_d(L-1) \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ b_M \end{bmatrix}.$$

Shank's method and Prony's method give good approximations when the number of poles and zeros of the designed filter equals or exceeds the number of poles and zeros of the actual filter. However, in practice only a sequence of data points is provided. Thus, an effective approach in using the above methods is to try different values of M and N until the desired frequency response is approximated within acceptable range of error [14]. Actually, this is a cumbersome approach. It is clear from the above discussion that these methods are not effective when approximating a high order IIR filter by a

lower order one. Finally, the designs obtained using the above methods are not always stable.

2.3 The CF (Caratheodory-Fejer) Method

The CF method is based on an extension due primarily to Takagi of a classical theorem in complex analysis of Caratheodory and Fejer [15], hence the name CF. This technique starts from a truncated or windowed impulse response to compute a stable, near optimal in the Chebyshev sense and uniformly weighted rational approximation to the complex transfer function $H(e^{j\omega})$.

2.3.1 Problem Description

The basis of the problem solved by this method is that having $H_K(z)$, where $H_K(z) = \sum_{n=0}^{K} h_k(n)z^{-n}$, find a rational transfer function

$$R_{MN}(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}} \qquad (2.25)$$

which approximates the function $H_K(z)$ on the unit circle with all poles inside the unit circle and normalized with $a_0=1$. We denote the set of all such functions by R_{MN}. Actually, an optimal rational approximation to $H_K(z)$ on the unit circle under Chebyshev norm exists but it is very difficult to compute such a function [15]. However, it happens that it is easy to determine the best

Chebyshev approximation $\tilde{R}^{*}_{MN}$ out of the larger class $\tilde{R}_{MN}$ of functions which are of the form

$$\tilde{R}_{MN} = \frac{\tilde{B}(z)}{A(z)} = \frac{\sum_{k=-\infty}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}} \quad . \tag{2.26}$$

This class of functions includes a noncausal impulse response components. The CF method computes this extended best approximation and truncates it to find the CF approximant $R^{(CF)}_{MN} \in R_{MN}$. Although the designed filter is no more optimal because of truncation of $\tilde{R}^{*}_{MN}$, it is exceedingly close to optimal in the Chebyshev sense [15].

Next, the theorem which the CF method is based on is presented. It also explains to us how to find $\tilde{R}^{*}_{MN}$.

Theorem 2.1: [19]

H_K has a unique best Chebyshev approximation $\tilde{R}^{*}_{MN}$ out of $\tilde{R}_{MN}$ and the error function $(H_K - \tilde{R}^{*}_{MN})(z)$ is an all-pass filter having constant modulus on $|z| = 1$. The error modulus is equal to the N-th singular value of the Hankel matrix $H_{v,K}$, i.e.,

$$\left\| H_K - \tilde{R}^{*}_{MN} \right\|_{\infty} = \sigma_N \tag{2.27}$$

where $\nu = M-N+1$ and $\sigma_N = 0$ for $N > K-\nu$. $\tilde{R}^{*}_{MN}$ is given by

$$\tilde{R}^{*}_{MN}(z) = H_K(z) - \sigma_N z^{-\nu} \frac{U_N(z)}{V_N(z^{-1})} \tag{2.28}$$

where $U_N(z)$ and $V_N(z)$ are formed from the Nth singular vectors of $H_{v,K}$ as

$$U_N(z) = \sum_{n=0}^{K-\nu} u_N(n) z^{-n} \tag{2.29}$$

$$V_N(z) = \sum_{n=0}^{K-\nu} v_N(z) z^{-n}. \tag{2.30}$$

Note that the theorem makes use of the singular value decomposition of the Hankel matrix $H_{\nu,K}$.

2.3.2 Problem Formulation

Given a finite-length impulse response $h_k(n)$, obtain $R_{MN}^{(CF)}$ which represents the causal part of the best Chebyshev approximation given in eq.(2.28) in the above theorem.

The following algorithm shows us how to find $R_{MN}^{(CF)}$.

2.3.3 The CF Algorithm

Starting from $h_k(n)$, which is assumed to be real for simplicity, the algorithm is as follows:

1) Set up the Hankel matrix $H_{\nu,K}$ and compute its eigenvalue λ_N which is the largest N+1 in modulus.

$$H_{\nu,K} = \begin{bmatrix} h_K(\nu) & h_K(\nu+1) & . & . & h_K(K) \\ h_K(\nu+1) & . & & . & 0 \\ . & . & & 0 & . \\ . & . & & . & . \\ h_K(K) & 0 & . & . & 0 \end{bmatrix}$$

where ν and K were defined above and the Hankel matrix entry (i,j) is defined as $h_K(i+j+\nu)$ with (i,j=0,1,2,.....) . For $i+j+\nu < 0$, $h(i+j+\nu) = 0$.

2) Compute the eigenvector V_N (eq.2.30) corresponding to λ_N. Note that we don't need to compute σ_N nor U_N because $h_K(n)$ is assumed to be real.

3) Evaluate the frequency response of the optimal (noncausal) Chebyshev approximation at L >> M+N+1 equally spaced points along the unit circle

$$\tilde{R}^{*}_{MN}(e^{j\omega_k}) = H_K(e^{j\omega_k}) - \lambda_N e^{-jv\omega_k}\frac{V_N(e^{j\omega_k})}{V_N(e^{-j\omega_k})}$$

$$\omega_k = \frac{2\pi k}{L} \quad , k = 0, 1, \ldots L-1.$$

4) Inverse Fourier transform $\tilde{R}^{*}_{MN}$ to obtain the impulse response of the extended rational Chebyshev approximation

$$\tilde{r}^{*}_{MN}(n) = FFT^{-1}(\tilde{R}^{*}_{MN}) = \frac{1}{L}\sum_{k=0}^{L}\tilde{R}^{*}_{MN}(e^{j\omega_k})\, e^{j\omega_k} \quad .$$

The first L/2 samples , n = 0,1,...,L/2 -1 , corresponds to the causal part.

Now , *for* $v \geq 0$ $(M \geq N-1)$ *:*

5) Window $\tilde{r}^{*}_{MN}$ to obtain the impulse response of the Hankel norm approximation

$$r^{(CF)}_{MN} = \begin{cases} \tilde{r}^{*}_{MN} & n=0,1,\ldots,\frac{L}{2}-1 \\ 0 & n=\frac{L}{2},\ldots,L-1 \end{cases}$$

6) Convert the nonparametric impulse response $r^{(CF)}_{MN}$ to parametric form $[a_i, b_j]$, i=1,...,N , j=1,...,M via Prony's method.

For $v < 0$ *:*

5) Window $\tilde{r}^{*}_{MN}$ as follows :

$$\hat{r}^{(CF)}_{N-1,N} = \begin{cases} \tilde{r}^{*}_{MN}(n+v) & n=0,1,\ldots,\frac{L}{2}-1 \\ 0 & n=\frac{L}{2},\ldots,L-1 \end{cases}$$

6) Convert $\hat{r}^{(CF)}_{N-1,N}$ to parametric form $[c_j, a_i]$ with i=1,...,N , j=1,..., N-1 via Prony's method and set $b_j=c_{j-v}$, j=0,...,M.

There are some important points to clarify about this method. First, this method gives a stable and near optimal in the Chebyshev sense rational approximation. Moreover, it is in fact optimal in the Hankel norm sense when $M \geq N-1$ [15]. Also, the CF method differs from model order reduction methods in two main aspects. First, it eliminates the restriction that M=N-1 as will be seen later in the optimal Hankel method. Thus, it has the flexibility to design IIR filters with any given number of poles and zeros. Second, it avoids using partial fraction expansion to obtain the causal part of the function since it could limit the length of the impulse response used. Actually, this length represents the size of the polynomial to be factored. Instead, it uses a spectral factorization technique based on FFT. Steps 3-6 in the algorithm represents this idea.

On the other hand, there is no choice over the frequency error measure $|\lambda_N|$ since the filter order (M,N) is prescribed. However, an alternative is to prescribe only the difference between the number of poles and zeros which is designated as *v* in the algorithm above and then decide on the final order of the filter after the eigenvalues of $H_{v,K}$ have been inspected [15].

CHAPTER III

SYSTEM BALANCING APPROXIMATION METHODS

System balancing approximation methods are state space approach based. They are totally different from the above methods. They deal with certain parameters related to system and control theory to search for lower order models which approximate high order systems efficiently. In the following sections, some of these methods will be presented.

3.1 Problem Description

Many real-life systems such as power system plants, distillation towers, and space shuttles are usually of a very high order and they need large state space models to be realized. This situation is unpleasant since high order models are difficult to analyze and simulate. Also, design methods based on them need a large amount of computation. Thus, it is highly favorable to get reduced-order models which approximate high order systems efficiently. In other words, we need to find a strong or a dominant subsystem of the original system. However, on which basis one can judge weather that the reduced-order model is a strong subsystem or not ?. One way to judge is to consider the controllability and observability of the states of the system simultaneously.

A strong subsystem corresponds to the most controllable and observable part of the original system [32]. This subsystem was first found by Moore [22] for continuous time systems. A discrete time version of it is found in [32] and it is discussed below.

A digital filter could be looked at as a discrete system which has a state variable representation. This enables us to apply model reduction techniques to digital filter approximation and design.

3.2 Internal Balanced Approximation Method

Consider the following minimal representation (A,B,C) of order n for a given stable discrete system (which could be an IIR digital filter) described by:

$$x(k+1) = A\ x(k) + B\ u(k) \tag{3.1}$$

$$y(k) = C\ x(k)$$

with:

$$W_c = \sum_{k=0}^{\infty} A^k B B^T (A^k)^T \tag{3.2}$$

$$W_o = \sum_{k=0}^{\infty} (A^k)^T C^T C A^k \tag{3.3}$$

where W_c and W_o are respectively the controllability and observability grammians of the system. Moore in [28] suggested an internal balanced representation $(\hat{A}, \hat{B}, \hat{C})_n$ where the state variables x(k) are ordered according to their

input-output influence and responsiveness. This balanced realization is obtained by applying a nonsingular transformation matrix T such that:

$$\hat{A}= TAT^{-1},\ \hat{B}= TB,\ \hat{C}= CT^{-1} \tag{3.4}$$

with:

$$T = \Sigma^{\frac{1}{2}}\ \hat{U}_o^T\ \Sigma_c^{-\frac{1}{2}}\ U_c^T\ . \tag{3.5}$$

We will explain later how to find this T.

The key idea to obtain such a transformation is to consider the product of W_c and W_o. The eigenvalues Σ^2 of this product, call it Δ, is invariant under a general (not necessarily T defined above) nonsingular state space transformation T :

$$\hat{\Delta} = \hat{W}_o\ \hat{W}_c = T^{-T}W_oW_cT^T = T^{-T}\Delta T^T \tag{3.6}$$

where $\hat{W}_o = T^{-T}W_oT^{-1}$, $\hat{W}_c = TW_cT^T$.

This invariance of Σ^2 enables us to find a state space transformation with the property that $\hat{W}_o = \hat{W}_c = \Sigma$ which is proved in [22] and [31]. This means that the controllability and observability grammians of the balanced system are equal and diagonal. A system with this property is as controllable as it is observable.

As mentioned above, the states of the balanced system are ordered according to their degree of controllability and observability. Thus, it is logical to think that if we kept the strongly controllable and observable states and through the weakly controllable and observable states, a suitable reduced

model would be obtained. Based on the idea of system balancing, an algorithm to find a reduced model for a given system is explained in the next section.

3.2.1 Balanced Approximation Algorithm [32]

Assume $(\hat{A},\hat{B},\hat{C})_n$ is a balanced system with Σ partitioned as follows

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{3.7}$$

where $\Sigma_1 = \text{diag}(\sigma_1,...,\sigma_r)$, $\Sigma_2 = \text{diag}(\sigma_{r+1},...,\sigma_n)$ and r is the order of the reduced model designed. Accordingly, the system will be partitioned as follows:

$$\hat{A}=\begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} \quad , \hat{B}=\begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} \quad , \hat{C}=\begin{bmatrix} \hat{C}_1 & \hat{C}_2 \end{bmatrix} . \tag{3.8}$$

If $\sigma_r >> \sigma_{r+1}$, it is logical to think of the reduced model $(\hat{A}_{11},\hat{B}_1,\hat{C}_1)_r$ as being the strongly controllable and observable part of the system. Thus, the above reduced model is considered as a suitable r-th order approximant to the original system. Moreover, it can be shown that if $(\hat{A},\hat{B},\hat{C})_n$ is asymptotically stable, then the reduced model $(\hat{A}_{11},\hat{B}_1,\hat{C}_1)_r$ is asymptotically stable [24]. The steps to find a balanced system along with its r-th order reduced model are as follows:

1) Solve for the grammians W_c and W_o as the unique positive definite solutions of the Lyapunov equations :

$$AW_cA^T - W_c = -B\,B^T \tag{3.9}$$

$$A^TW_oA - W_o = -C^T\,C\,. \tag{3.10}$$

2) Perform the eigenvalue decomposition of W_c :

$$W_c = U_c\Sigma U_c^T \quad , \quad U_cU_c^T = I. \tag{3.11}$$

3) Form the matrix $\hat{W}_o$

$$\hat{W}_o = \Sigma_c^{\frac{1}{2}}U_c^TW_oU_c\Sigma_c^{\frac{1}{2}}. \tag{3.12}$$

4) Compute the eigenvalue decomposition of $\hat{W}_o$:

$$\hat{W}_o = \hat{U}_o\,\Sigma^2\,\hat{U}_o^T \quad , \quad \hat{U}_o\,\hat{U}_o^T = I\,. \tag{3.13}$$

5) The balanced transformation is then :

$$T = \Sigma^{\frac{1}{2}}\,\hat{U}_o^T\,\Sigma_c^{-\frac{1}{2}}U_c^T\,. \tag{3.14}$$

6) Compute the balanced realization $(\hat{A},\hat{B},\hat{C})$ as:

$$\hat{A} = TAT^{-1} \quad , \quad \hat{B} = TB \quad , \quad \hat{C} = CT^{-1}\,. \tag{3.15}$$

7) The reduced model is the top left corner subsystem $(\hat{A}_{11}\,,\hat{B}_1\,,\hat{C}_1)$ of $(\hat{A},\hat{B},\hat{C})$ of order r.

As a matter of fact, there are many algorithms developed to obtain the above balanced model. A highly efficient algorithm was introduced by Laub in [51].

3.3 Suboptimal Hankel Approximation Methods

The balanced realization discussed in the last section can be also obtained using a totally different approach. It depends on the singular value decomposition of the Hankel matrix formed from the impulse response of a given system. In the subsections to follow, two Hankel methods are discussed: Kung's method and Kimura's method.

3.3.1 Kung Method

This method requires only the impulse response (Markov parameters) given by :

$$h(k) = C\,A^{k-1}\,B \qquad , \; k = 1,2,3,... \tag{3.16}$$

where $(A,B,C)_n$ is a realization of the system under consideration. If the system is asymptotically stable, i.e. $|\lambda_i(A)| < 1$ for all i, then the infinite Hankel matrix H is given by :

$$H = \begin{bmatrix} h(1) & h(2) & h(3) & . & . \\ h(2) & h(3) & . & . & . \\ h(3) & . & & & \\ . & . & & & \\ . & . & & & \end{bmatrix} = OG \tag{3.17}$$

where O and G are the infinite observability and controllability matrices of $(A,B,C)_n$. Also, the observability and controllability grammians can then be expressed as [32]:

$$W_o = O^T O \quad , \quad W_c = GG^T \tag{3.18}$$

Before discussing how to obtain a balanced realization via Hankel factorization, two basic results from realization theory are considered. The first result is that rank(H) = n. This implies that although H is infinite, it has exactly n nonzero singular values. This result enables us to find an exact finite realization of H. The second result explains the relation between the i-th eigenvalue of $\Delta = W_c W_o$ which is a finite dimensional matrix of order n and the i-th singular value of H as follows: [31]

$$\begin{aligned} \lambda_i[\Delta] &= \lambda_i[GG^T O^T O] \\ &= \lambda_i[OGG^T O^T] \\ &= \lambda_i[HH^T] \\ &= \sigma_i^2[H] \end{aligned} \tag{3.19}$$

where $1 \leq i \leq n$. Thus, the i-th eigenvalue of Δ equals the square of the i-th singular value of H. The impact of this result is that although H is infinite, its singular values can still be computed by utilizing eq.(3.19).

3.3.1.1 Balanced realization via Hankel factorization. [32] It is well known that every factorization of H leads to some state space realization [52]. SVD of H is utilized to obtain a specific factorization from which the balanced realization is reached. SVD of H is given by

$$H = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{3.20}$$

where Σ is defined in eq.(3.7). From eq.(3.20), H can be factored as follows:

$$H = OG \quad , \quad G = \Sigma^{\frac{1}{2}} V^T \quad , \quad O = U\Sigma^{\frac{1}{2}}. \tag{3.21}$$

Then:

$$\hat{W}_c = GG^T = \Sigma^{\frac{1}{2}} V^T V \Sigma^{\frac{1}{2}} = \Sigma \tag{3.22}$$

$$\hat{W}_o = O^T O = \Sigma^{\frac{1}{2}} U^T U \Sigma^{\frac{1}{2}} = \Sigma. \tag{3.23}$$

This is actually a consequence of the second result stated above. The above factorization corresponds to the following balanced realization which is obtained as a special case of Ho's algorithm: [32,52]

$$\hat{A} = \Sigma^{-\frac{1}{2}} U^T \bar{H} V \Sigma^{-\frac{1}{2}} \tag{3.24a}$$

$$\hat{B} = \text{first m columns of } \Sigma^{\frac{1}{2}} V^T \tag{3.24b}$$

$$\hat{C} = \text{first p rows of } U\Sigma^{\frac{1}{2}} \tag{3.24c}$$

where $\bar{H}$ is a shifted version of H. Now, let us consider how to obtain the above realization. For $\hat{B}$ and $\hat{C}$, they are clearly obtained from the definition of controllability and observability where:

$$G = \begin{bmatrix} B & AB & \dots \end{bmatrix} \tag{3.25}$$

$$O^T = \begin{bmatrix} C & CA & \dots \end{bmatrix}. \tag{3.26}$$

For $\hat{A}$, we need to define the following matrices :

G_b, O_b : controllability and observability of the balanced realization respectively.

$O_b^{\uparrow}$: submatrix of O_b shifted up p rows from O_b.

$G_b^{\leftarrow}$: submatrix of G_b shifted m columns to the left from G_b.

O_b^{+} : pseudo-inverse of O_b , $O_b^{+} = \Sigma^{-\frac{1}{2}} U^T$.

G_b^{+} : pseudo-inverse of G_b , $G_b^{+} = V\Sigma^{-\frac{1}{2}}$.

From the shift property of the Hankel matrix [32], the following equation is obtained:

$$\bar{H} = \begin{bmatrix} \hat{C}\hat{A}\hat{B} & \hat{C}\hat{A}^2\hat{B} & \cdots \\ \hat{C}\hat{A}^2\hat{B} & & \\ \vdots & & \end{bmatrix} = \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ \vdots \end{bmatrix} \hat{A} \begin{bmatrix} \hat{B} & \hat{A}\hat{B} & \cdots \end{bmatrix}$$

$$= O_b \hat{A} G_b = O_b^{\uparrow} G_b = O_b G_b^{\leftarrow}. \qquad (3.27)$$

Thus, $\hat{A}$ is the unique solution to any of the three identities:

$$\bar{H} = O_b \hat{A} G_b \qquad (3.28a)$$

$$O_b^{\uparrow} = O_b \hat{A} \qquad (3.28b)$$

$$G_b^{\leftarrow} = \hat{A} G \qquad (3.28c)$$

and $\hat{A}$ is given by any of these equations corresponding to the above identities respectively:

$$\hat{A} = O_b^{+} \bar{H} G_b \qquad (3.29a)$$

$$= O_b^{+} O_b^{\uparrow} \qquad (3.29b)$$

$$= G_b^{\leftarrow} G_b. \qquad (3.29c)$$

3.3.1.2 Reduced Model. Following a similar algorithm to the one described above, Kung in [27] developed an r-th order reduced model which is given by:

$$\hat{A}_{11} = O_1^{\dagger}\overleftarrow{O_1} = \Sigma_1^{-\frac{1}{2}}U_1^T(U_1\Sigma_1^{\frac{1}{2}}) \tag{3.30a}$$

$$\hat{B}_1 = \text{first m columns of } G_1 = \Sigma_1^{\frac{1}{2}}V_1^T \tag{3.30b}$$

$$\hat{C}_1 = \text{first p rows of } O_1 = U_1\Sigma_1^{\frac{1}{2}}. \tag{3.30c}$$

Note that there is no need to find the balanced realization in this case. Only SVD of H is required. In section 3.3.2, we shall show how to derive the above equations when Kung's method is applied to an IIR digital filter approximation.

A very important thing to note is that the reduced model obtained using this method is equivalent to the reduced model obtained by balanced approximation (Moore method). This is stated in the following lemma.

Lemma 3.1: [31]

Let (A_{11}^*, B_1^*, C_1^*) be the reduced model obtained by Kung's method from the balanced system $(\hat{A}, \hat{B}, \hat{C})_n$ which is partitioned as eq.(3.8) where $\hat{A}_{11}$ is r x r , $\hat{B}_1$ is r x m and $\hat{C}_1$ is p x r. Then :

$$\hat{A}_{11} = A_{11}^* \quad , \quad \hat{B}_1 = B_1^* \quad , \quad \hat{C}_1 = C_1^*. \tag{3.31}$$

3.3.2 Kimura and Honoki Method [42]

In this method, suboptimal Hankel approach is applied to FIR filters with LPHC to obtain a lower order IIR filter with a nearly LPHC. Note that h(n) is finite in this case and hence also the Hankel matrix.

Let $(A,B,C, h_0)_n$ be a minimal realization of a stable transfer function h(z) of order n, i.e.

$$h(z) = h_0 + C\,(zI - A)^{-1}\,B \tag{3.33}$$

$$= h_0 + h_1 z^{-1} + h_2 z^{-2} + \ldots + h_{N-1} z^{-(N-1)} \tag{3.34}$$

and the Hankel matrix of h(z) is

$$H = \begin{bmatrix} h_1 & h_2 & . & . & h_{N-1} \\ h_2 & h_3 & . & & 0 \\ . & & & & . \\ . & & & & . \\ h_{N-1} & 0 & . & . & 0 \end{bmatrix}. \tag{3.35}$$

The observability and controllability of the above realization is given respectively by :

$$O = \begin{bmatrix} C \\ CA \\ . \\ . \\ CA^{N-2} \end{bmatrix}, \quad G = \begin{bmatrix} B & AB & . & . & A^{N-2}B \end{bmatrix}. \tag{3.36}$$

The first step to obtain a reduced IIR model is to perform SVD to H which is given by eq.(3.20). Then, set the observability and controllability of the balanced system to be respectively:

$$O_b = \begin{bmatrix} \hat{B} & \hat{A}\hat{B} & . & . & \hat{A}^{N-2} & \hat{B} \end{bmatrix} = \begin{bmatrix} \Sigma_1^{\frac{1}{2}} V_1 \\ \Sigma_2^{\frac{1}{2}} V_2 \end{bmatrix} \tag{3.37}$$

$$G_b = \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ . \\ . \\ \hat{C}\hat{A}^{N-2} \end{bmatrix} = \begin{bmatrix} U_1\Sigma_1^{\frac{1}{2}} & U_2\Sigma_1^{\frac{1}{2}} \end{bmatrix} . \tag{3.38}$$

Since $O_b \hat{A} = \begin{bmatrix} \hat{C}\hat{A} \\ . \\ . \\ . \\ \hat{C}\hat{A}^{N-1} \end{bmatrix}$ is one-row shift upward of O_b, we have:

$$\begin{bmatrix} U_1\Sigma_1^{\frac{1}{2}} & U_2\Sigma_2^{\frac{1}{2}} \end{bmatrix} \hat{A} = \begin{bmatrix} U_1\Sigma_1^{\frac{1}{2}} & U_2\Sigma_2^{\frac{1}{2}} \end{bmatrix}^{\uparrow} \tag{3.39}$$

or in partitioned form:

$$U_1\Sigma_1^{\frac{1}{2}} \hat{A}_{11} + U_2\Sigma_2^{\frac{1}{2}} \hat{A}_{21} = (U_1\Sigma_1^{\frac{1}{2}}) \tag{3.40a}$$

$$U_1\Sigma_1^{\frac{1}{2}} \hat{A}_{12} + U_2\Sigma_2^{\frac{1}{2}} \hat{A}_{22} = (U_2\Sigma_2^{\frac{1}{2}}) . \tag{3.40b}$$

If equation (3.40a) is premultiplied by U_1^T, then the reduced IIR model is given by:

$$\hat{A}_{11} = \Sigma_1^{-\frac{1}{2}} U_1^T (U_1\Sigma_1^{\frac{1}{2}}) \tag{3.41a}$$

$$\hat{B}_1 = \text{The first column of } \Sigma_1^{\frac{1}{2}} V_1 \tag{3.41b}$$

$$\hat{C}_1 = \text{The first row of } U_1\Sigma_1^{\frac{1}{2}}. \tag{3.41c}$$

It is clear from the above formulae that there is no need to find the balanced realization of the system. Instead, SVD of the Hankel matrix is required only. Note that $(\hat{A}_{11}, \hat{B}_1, \hat{C}_1, h_0)$ is not necessarily balanced but the " ^ " notation follows from the original notation of the balanced realization.

Actually, Kimura's method is basically Kung's method applied to a single-input single-ouput system with the D-term which is neglected in Kung's method forced to be equal to h_0 in Kimura's method. The effect of the D-term will be investigated more in the remarks in section 3.5.

3.4 Minimum Sensitivity Method

One of the main advantages of balanced realization that it has the minimum sensitivity property with respect to parameter variations and it can simultaneously attain minimum sensitivity to both parameter variation and roundoff noise by applying a specific transformation [23,35,36,37]. However, the reduced models obtained using the previous methods are not balanced. This problem is handled by Minimum sensitivity method (MS method) where it finds a reduced model which is also balanced. In this section, MS filter design with respect to parameter variation is introduced first and then with respect to both parameter variation and roundoff noise.

3.4.1 MS Filter Design with Respect to Parameter Variation

As mentioned above, a balanced realization has the minimum sensitivity property to parameter variations. However, the reduced order model $(\hat{A}_{11}, \hat{B}_1, \hat{C}_1)_r$ obtained in the previous sections is not in general balanced. Thus, the problem handled here is to find a reduced order model which is still balanced.

Assume $(\hat{A},\hat{B},\hat{C},D)_n$ is a balanced system and it is partitioned as follows:

$$\hat{A}=\begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}, \quad \hat{B}=\begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix}, \quad \hat{C}=\begin{bmatrix} \hat{C}_1 & \hat{C}_2 \end{bmatrix}$$

with $$\Sigma=\begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}.$$

Accordingly, the state space realization can be written as follows:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix}=\begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}+\begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix}u(k) \qquad (3.42a)$$

$$y(k)=\begin{bmatrix} \hat{C}_1 & \hat{C}_2 \end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}+D\,u(k) \qquad (3.42b)$$

or:

$$x_1(k+1)=\hat{A}_{11}\,x_1(k)+\hat{A}_{12}\,x_2(k)+\hat{B}_1\,u(k) \qquad (3.43a)$$

$$x_2(k+1)=\hat{A}_{21}\,x_1(k)+\hat{A}_{22}\,x_2(k)+\hat{B}_2\,u(k) \qquad (3.43b)$$

$$y(k)=\hat{C}_1\,x_1(k)+\hat{C}_2\,x_2(k)+D\,u(k)\,. \qquad (3.43c)$$

Since $x_2(k)$ reaches steady state more quickly than $x_1(k)$, the effect of $x_2(k)$ on the system is less significant than that of $x_1(k)$ [53]. But, this is valid only for 2-dynamics or singularly perturbed systems [54]. In this situation, with the approximation $x_2(k+1) \cong x_2(k)$, eq.(3.44b) is solved to get $x_2(k)$:

$$(I-\hat{A}_{22})x_2(k)=\hat{A}_{21}\,x_1(k)+\hat{B}_2\,u(k) \qquad (3.44)$$

Thus,

$$x_2(k)=(I-\hat{A}_{22})^{-1}\hat{A}_{21}\,x_1(k)+(I-\hat{A}_{22})^{-1}\hat{B}_2\,u(k)\,. \qquad (3.45)$$

It can be proved that since $(\hat{A}, \hat{B}, \hat{C}, \hat{D})_n$ is balanced, it is asymptotically stable. This insures the existence of the matrix $(I-\hat{A}_{22})^{-1}$. By substituting eq.(3.45) into eq.(3.43a) and eq.(3.43c), the following is obtained:

$$x_1(k+1) = [\hat{A}_{11} + \hat{A}_{12}(I-\hat{A}_{22})^{-1}\hat{A}_{21}]\,x_1(k) + [\hat{B}_1 + \hat{A}_{12}(I-\hat{A}_{22})^{-1}\hat{B}_2]\,u(k) \tag{3.46}$$

$$y(k) = [\hat{C}_1 + \hat{C}_2(I-\hat{A}_{22})^{-1}\hat{A}_{21}]\,x_1(k) + \hat{C}_2(I-\hat{A}_{22})^{-1}\hat{B}_2\,u(k)\,. \tag{3.47}$$

The reduced order balanced filter is defined as:

$$x^*(k+1) = A^*\,x^*(k) + B^*\,u(k) \tag{3.48a}$$

$$y^*(k) = C^*\,x^*(k) + D^*\,u(k) \tag{3.48b}$$

where

$$A^* = \hat{A}_{11} + \hat{A}_{12}(I-\hat{A}_{22})^{-1}\hat{A}_{21} \tag{3.49a}$$

$$B^* = \hat{B}_1 + \hat{A}_{12}(I-\hat{A}_{22})^{-1}\hat{B}_2 \tag{3.49b}$$

$$C^* = \hat{C}_1 + \hat{C}_2(I-\hat{A}_{22})^{-1}\hat{A}_2 \tag{3.49c}$$

$$D^* = D + \hat{C}_2(I-\hat{A}_{22})^{-1}\hat{B}_2\,. \tag{3.49d}$$

The above defined reduced realization is balanced and asymptotically stable. This is stated in the following theorem [55].

Theorem 3.1:

i) The reduced order filter $(A^*, B^*, C^*, D^*)_r$ satisfies

$$A^*\Sigma_1 A^{*T} - \Sigma_1 + B^*B^{*T} = 0 \tag{3.50}$$

$$A^{*T}\Sigma_1 A^* - \Sigma_1 + C^{*T}C^* = 0 \tag{3.51}$$

ii) If Σ_1 and Σ_2 have no diagonal entries in common, then the above r-th order approximation is asymptotically stable.

3.4.2 MS Filter Design with Respect to both Parameter Variation and Roundoff Noise

The above reduced-order filter has the minimum sensitivity property to parameter variation only. However, it has been shown [36,37] that if the observability grammian matrix and controllability grammian matrix are in the following form:

$$M = \rho^2 W \tag{3.52}$$

where

$$\rho = \sum_{i=1}^{r} \frac{\sigma_i}{r} \tag{3.53}$$

with r representing the order of the filter and W is given by

$$W = \begin{bmatrix} 1 & & & \\ & 1 & X & \\ & & . & \\ & X & & . \\ & & & 1 \end{bmatrix} \tag{3.54}$$

where the matrix X is generally not equal to zero, then the filter will simultaneously achieve the minimum sensitivity property to parameter variation and roundoff noise. This form of W_c and W_o is obtained via a nonsingular transformation matrix $\bar{T}$. The new state space realization will be

$$(\bar{A}, \bar{B}, \bar{C}, \bar{D})_r = (\bar{T}^{-1} A^* \bar{T}, \bar{T}^{-1} B^*, C^* \bar{T}, D^*)_r \,. \tag{3.55}$$

The nonsingular matrix $\bar{T}$ is chosen as

$$\bar{T} = \delta P^T \tag{3.56}$$

with $$PP^T = I \tag{3.57}$$

where $$\delta = \left[\sum_{i=1}^{r} \frac{\sigma_i}{r} \right]^{\frac{1}{2}} \tag{3.58}$$

and $$P = P_r P_{r-1} \ldots P_i \ldots P_2 \tag{3.59}$$

with

$$P_i = \begin{bmatrix} I & . & 0 & . & 0 \\ . & \cos\psi_i & . & \sin\psi_i & . \\ 0 & . & I & . & 0 \\ . & -\sin\psi_i & . & \cos\psi_i & . \\ 0 & . & 0 & . & I \end{bmatrix} \qquad i = 2,\ldots,r. \tag{3.60}$$

Given $(A, B, C)_r$, The following steps explain how to find P: [56]

1) Form the covariance matrix K_0 and the unit noise matrix W_0 given respectively by:

$$K_0 = \sum_{m=0}^{\infty} A^m B B^T (A^m) \tag{3.61}$$

$$W_0 = \sum_{m=0}^{\infty} (A^m)^T C^T C A \, . \tag{3.62}$$

Both K_0 and W_0 become stable after a finite number of terms are computed. Thus, there is mostly no problem from computational point of view.

2) Form the matrix $\Phi_0 = K_0 W_0$ and obtain its eigenvalues denoted by θ_i^2.

3) Form the following diagonal matrix :

$$\Lambda = \begin{bmatrix} \mu_1^2 & & 0 \\ & \mu_2^2 & \\ & & . \\ 0 & & \mu_r^2 \end{bmatrix} \tag{3.63}$$

where $$\mu_i^2 = \frac{r\theta_i}{\sum_{m=1}^{r} \theta_m} \quad . \tag{3.64}$$

4) Obtain P_2 such that the first diagonal element of the following product is equal to unity:

$$M_1 = P_2 \Lambda P_2^T = \begin{bmatrix} \cos\psi_2 & \sin\psi_2 & \\ -\sin\psi_2 & \cos\psi_2 & 0 \\ & & \\ 0 & & I \end{bmatrix} \begin{bmatrix} \mu_1^2 & 0 & \\ 0 & \mu_2^2 & 0 \\ & & \cdot \\ 0 & & \mu_r^2 \end{bmatrix} \begin{bmatrix} \cos\psi_2 & -\sin\psi_2 & \\ \sin\psi_2 & \cos\psi_2 & 0 \\ & & \\ 0 & & I \end{bmatrix}$$

$$= \begin{bmatrix} \mu_1^2\cos^2\psi_2 + \mu_2^2\sin^2\psi_2 & (\mu_2^2 - \mu_1^2)\cos\psi_2\sin\psi_2 & & \\ (\mu_2^2 - \mu_1^2)\cos\psi_2\sin\psi_2 & \mu_2^2\cos^2\psi_2 + \mu_1^2\sin^2\psi_2 & & 0 \\ & & \mu_3^2 & \\ & & & \cdot \quad 0 \\ & & & \cdot \\ 0 & & 0 & \mu_r^2 \end{bmatrix}. \tag{3.65}$$

If $\cos\psi_2$ and $\sin\psi_2$ is chosen as

$$\cos\psi_2 = \left(\frac{\mu_2^2 - 1}{\mu_2^2 - \mu_1^2}\right)^{\frac{1}{2}} \quad , \quad \sin\psi_2 = \left(\frac{1 - \mu_2^2}{\mu_2^2 - \mu_1^2}\right)^{\frac{1}{2}} , \tag{3.66}$$

then M_1 becomes

$$M_1 = \begin{bmatrix} 1 & \sqrt{(1-\mu_1^2)(\mu_2^2 - 1)} & & 0 \\ \sqrt{(1-\mu_1^2)(\mu_2^2 - 1)} & \mu_1^2 + \mu_2^2 - 1 & & \\ & & \mu_3^2 & 0 \\ & & & \cdot \\ 0 & & & \cdot \\ & & 0 & \mu_r^2 \end{bmatrix} \tag{3.67}$$

5) Substitute eq.(3.66) in eq.(3.65) to obtain P_2.

6) Repeat the above steps to obtain P_3 , .., P_i, ..,P_r with

$$M_i = P_{i+1} M_{i-1} P_{i+1}^T$$

being diagonal.

7) Apply eq.(3.59) to calculate P.

3.5 Remarks

The following remarks cover some important aspects regarding the above methods.

1) The balancing methods discussed above are not optimal although they compare favorably with other methods. The following theorem explains the reason for this non-optimality [32].

Theorem 3.2:

Let

$$P = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

be a rank n matrix, where $\Sigma_1 = \text{diag}(\sigma_1, ..., \sigma_r)$ and $\Sigma_2 = \text{diag}(\sigma_{r+1}, ..., \sigma_n)$ with $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n \geq 0$. Then, a best rank r approximation in the spectral norm sense to P is:

$$Q = U_1 \Sigma_1 V_1^T .$$

Moreover, the error is:

$$E = \|P - Q\|_s = \sigma_{r+1} .$$

However, Q is not a Hankel matrix in general. Thus, it doesn't have an exact realization. Specifically, there is no exact solution for the A matrix. The above balancing methods find an approximate realization for Q, hence it loses optimality. Actually, there exists an optimal rank r approximation which is also Hankel as will be shown later.

2) The approximation error of the frequency response has an upper bound which is given by :

$$E_\infty = \|H(z) - H_r(z)\|_\infty \leq 2 \sum_{i=r+1}^{n} \sigma_i \ . \qquad (3.68)$$

and the strict inequality holds if $\sigma_i \neq \sigma_{i+1}$ for any i, $r \leq i \leq n-1$ [38]. H(z) and $H_r(z)$ are defined as:

$$H(z) = C(zI - A)^{-1}B \qquad (3.69)$$

$$H_r(z) = \hat{C}_1 (zI - \hat{A}_{11})^{-1} \hat{B}_1 \qquad (3.70)$$

Note that the D-term is not included in the frequency error bound. Actually, the D-term of the reduced order model is set to zero in both internal balanced approximation and Kung's method. Therefore, the frequency response approximation error for a system with a finite D-term may exceed the bound given above. In Kimura and Honoki method, the D-term is considered to be h_0. For MS method, an expression is obtained for it. The effect of the D-term will be investigated further in later sections. On the other hand, impulse response approximation [Hankel norm] is not affected by the D-term since h(i) is given by $CA^{i-1}B$ for $i > 0$. Actually, these methods are based on approximating the impulse response of a given system by the reduced model which explains the reason for neglecting the D-term.

3) The conditions for the controllability, observability, and stability of the reduced order model $(\hat{A}_{11}, \hat{B}_1, \hat{C}_1)_r$, which is obtained from the balanced system $(\hat{A}, \hat{B}, \hat{C})_n$, are stated in the following two lemma.

Lemma 3.2: [24,38]

Assume that the system (A, B, C) is asymptotically stable and that either the controllability grammian or observability grammian is nonsingular and diagonal. Then every subsystem is asymptotically stable.

Lemma 3.3: [24,38]

For the balanced system $(\hat{A}, \hat{B}, \hat{C})_n$, if $\sigma_r > \sigma_{r+1}$, then the subsystem of order r $(\hat{A}_{11}, \hat{B}_1, \hat{C}_1)_r$ is controllable and observable.

The first lemma shows that if the balanced realization is stable, then every subsystem of it is always stable. The second lemma gives a sufficient but not necessary condition for controllability and observability of the subsystem.

4) The balanced approximation developed by Moore might fail in retaining the most significant states for impulse response approximation [25,26]. An alternative criterion was suggested in [25] which is based on considering the contributions of the states to the impulse response norm. This alternative criterion can lead to a better reduced model. However, there are some cases where Moore's technique leads to a better approximation of the impulse response [57].

5) It is generally thought that if $\sigma_{r+1} << \sigma_r$, then the approximation would be good. However, good approximations were also obtained when the above condition is not satisfied, i.e. σ_{r+1} is close to σ_r. This could be seen from the error bound given by eq.(3.68) [33].

3.6 Optimal Hankel Methods

The balanced approximation techniques introduced above are not optimal. In this section two optimal methods which use the Hankel approach are introduced: optimal Hankel without a D-term (OPH) and optimal Hankel with a D-term (OPHD). Moreover, a solution is suggested to compensate for the D-term in OPH method.

3.6.1 OPH Method [32]

As mentioned above, a best rank r approximation of the Hankel matrix is not a Hankel matrix in general. In fact, the reduced model obtained using the balanced approximation techniques is a least square approximation of this best rank r approximation. However, Adamjan et al. [58] proved that there exists a rank r Hankel matrix which approximates optimally the infinite Hankel matrix generated by a scalar f(z) which will be defined later. Based on this result, Bettayeb [32] developed an optimal Hankel algorithm to approximate finite dimensional systems.

The optimal Hankel approximation algorithm is based on the following two theorems proved in [58]

Theorem 3.3: (time domain)

Let $\sigma_{r+1} < \sigma_r$ denote the (r+1)st singular value of a bounded Hankel matrix H. Then, there exists a unique Hankel matrix Λ^r of rank r which minimizes the spectral norm $\|H - \Lambda\|_s$ over all bounded Hankel matrices Λ of rank r

with an error bound given by $\|H-\Lambda^r\|_s = \sigma_{r+1}$. Moreover, $\Lambda^r = H - H(\phi^r(z))$

where

$$\phi^r(z) = \sigma_{r+1} \frac{\mu^{r+1}(z)}{\nu^{r+1}(z)} \quad . \tag{3.71}$$

μ^{r+1} and ν^{r+1} are defined respectively as follows:

$$\mu^{r+1}(z) = \sum_{j=1}^{\infty} u_j^{r+1} z^{-j}$$
$$\nu^{r+1}(z) = \sum_{j=1}^{\infty} v_j^{r+1} z^{j-1}$$

where μ^{r+1} and ν^{r+1} are the (r+1)th columns of U and V respectively with $H(f) = U\Sigma V^T$. u_j^{r+1} and v_j^{r+1} are the jth components of μ^{r+1} and ν^{r+1}.

Theorem 3.4: (frequency domain)

Let $f \in L_\infty$ and let $\sigma_{r+1} < \sigma_r$ denote the (r+1)st singular value of H(f). Then, there exists a unique function g^r which minimizes $\|f-g\|_\infty$ over the class of functions $g = g_s + g_u$ where g_s is a strictly proper rational function of degree r with poles in $|z| < 1$ and g_u is any function bounded on the unit circle $|z| = 1$ with $c_k(g_u) = 0$, $k > 0$. Moreover,

$$g^r(z) = f(z) - \phi^r(z) \tag{3.72}$$

and $\|f-g^r\|_\infty = \sigma_{r+1}$.

f(z) is an integrable function on the unit circle $|z| = 1$ and its Fourier coefficients are defined as:

$$c_k(f) = \frac{1}{2\pi} \int_0^{2\pi} z^k f(z) d\theta, \quad z = e^{j\theta} \tag{3.73}$$

Based on the above two theorems and on other results [31], an algorithm is suggested to calculate g_s^r for the special case in which f(z) represents the transfer function $C(zI - A)^{-1}B$ of a given system. Our interest here is the approximation of digital filters (a finite length data in general). Since f(z) can be expressed as a rational function [31], digital filters can be considered as a special case of f(z) if h_0 of the impulse response of the filter is neglected. Let

$$f(z) = \frac{n(z)}{d(z)}$$

and

$$H(z) = h_1 z^{-1} + h_2 z^{-2} + ... + h_n z^{-n} \ .$$

Then, H(z) could be written as

$$H(z) = \frac{h_1 z^{n-1} + h_2 z^{n-2} + ... + h_n}{z^n} = \frac{n(z)}{d(z)} \ . \qquad (3.74)$$

The steps to obtain the stable part g_s^r of the best r-th order approximation g^r in the Hankel norm sense for H(z) is listed below [32]. Moreover, an explicit expression is provided for g^r. It is assumed that real finite or truncated impulse response is given.

1) Form the finite Hankel matrix

$$H = \begin{bmatrix} h_1 & h_2 & . & . & h_n \\ h_2 & h_3 & & . & \\ . & & . & & \\ . & . & & 0 & \\ h_n & & & & \end{bmatrix}$$

2) Find the eigenvalues of H by $|H - \lambda I| = 0$.

3) Determine the reduced model order r.

4) Find the eigenvector M associated with the eigenvalue λ_{r+1}, where $M = (m(1), m(2), \ldots, m(n))^T$.

5) Form the transfer function

$$g^r(z) = \frac{\sum_{j=1}^{n-1} z^{n-j-1} \sum_{i=1}^{j} h(j-i+1)M(n-i+1)}{\sum_{i=1}^{n} M(i)z^{i-1}} \qquad (3.75)$$

6) Perform partial fraction expansion on $g^r(z)$.

7) Retain the stable terms. These terms will constitute g_s^r.

This method is optimal in Hankel norm sense (time domain). However, for L_∞ (frequency domain) it is not optimal unless r = n-1 [31]. The reason is that the optimal minimizing function g in theorem 3.4 is not stable in general. By throwing the unstable part g_u, the function is no more optimal. However, the case is different in time domain since the proper or stable part of g^r namely g_s^r is precisely of degree r, i.e:

$$H(g_s^r(z)) = H(g^r) = \Lambda^r \qquad (3.76)$$

which is the minimizing Hankel matrix given in theorem 3.3.

The frequency response error bound is given by the following inequality:

$$E_\infty = \|H(z) - g_s^r(z)\|_\infty = 2 \sum_{i=r+1}^{n} \sigma_i \qquad (3.77)$$

and when r = n-1, it becomes:

$$E_\infty = \|H(z) - g_s^r(z)\|_\infty = \sigma_n$$

which is optimal per theorem 3.4. For time domain, the error is always equal to σ_{r+1} with or without the unstable part of $g^r(z)$.

3.6.1.1 The D-term problem and a suggested solution. The remark about the D-term which was mentioned in the last section still holds for this method where the D-term is neglected. A possible solution to this problem is to assign h_0 to be the D-term of g_s^r. In this case, the Hankel norm is not affected since the Hankel matrix doesn't depend on the D-term. Also, the least squares error (LSE) of the impulse response given by eq.(2.3) will be less. For the frequency response, the error will be within the error bound given by eq.(3.77) since the D-term of the original model is canceled by the D-term of the reduced model (remark 2 in sec. 3.4). The effect of this modification on the performance of the reduced model will be shown later in the examples. OPH method with the D-term forced to it will be called " OPHd" to distinguish it from the next method.

3.6.2 OPHD Method

This method could be looked at as a generalization of the previous method. Glover in [33] derived characterizations of all solutions to the optimal Hankel norm multivariable approximation problem. Moreover, he explored the rule of the D-term which didn't receive much importance in the previous methods as mentioned previously. It is proved that the frequency response error or L_∞ norm error could be reduced to half by a proper choice of the D-term. This is stated in the following theorem [33].

Theorem 3.5:

Let $g_s^r(z)$ be an optimal Hankel norm approximation of degree r to the stable rational transfer function f(z). Also, let $g_u(z) \in H_-^\infty$ (unstable) be such that $g_s^r(z) + g_u(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}$. Then there exists D_0 such that

$$\|f(z) - g_s^r(z) - D_0\|_\infty \le \sigma_{r+1}(f(z)) + \sum_{1 \le i \le n-r-k} \sigma_{i+r+k}(f(z)) \qquad (3.78)$$

where k is the multiplicity of σ_{r+1}.

The state space representation $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ given in the above theorem is balanced and it represents an optimal approximation in both Hankel norm and Chebyshev norm. It consists of a stable function $g_s^r(z)$, which is an optimal Hankel approximation of degree r, and unstable or anticausal function $g_u(z)$.

When k=1, the right hand of eq.(3.78) becomes $\sum_{i=r+1}^{n} \sigma_i(f(z))$ which is half of the error bound provided in eq.(3.68). A proof of the above theorem can be found in [33]. On the other hand, the minimum achievable error is given by the following inequality:

$$\|f(z) - g_s^r(z)\|_\infty \ge \sigma_{r+1}(f(z)) \; . \qquad (3.79)$$

This error bound is achieved when an optimal anticausal approximation $g_u(z)$ is added to the optimal Hankel approximation $g_s^r(z)$ such that $\frac{f(z) - g_s^r(z) - g_u(z)}{\sigma_{r+1}(f(z))}$ is an all pass function in which case:

$$\|f(z) - g_s^r(z) - g_u(z)\|_\infty = \sigma_{r+1}(f(z)). \qquad (3.80)$$

In other words, the error function $E(z) = f(z) - g_s^r(z) - g_u(z)$ should be an all pass function to obtain an optimal approximation in Chebyshev sense.

An algorithm to find an optimal Hankel approximation with a D-term is outlined below [33]. It is assumed that the system to be approximated is real, stable and $\sigma_r > \sigma_{r+1} > \sigma_{r+2}$.

Algorithm:

1) It is assumed that a state-space realization of f(z), (A, B, C, D) has been supplied along with the degree of approximation , r.

2) Form a minimal balanced realization of f(z). Let the Hankel singular values be $\sigma_1 \geq \sigma_2 ... \geq \sigma_r > \sigma_{r+1} > \sigma_{r+2} .... \geq \sigma_n > 0$, and let the balanced realization of state dimension n be reordered so that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} , \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} , \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r, \sigma_{r+2}, ..., \sigma_n, \sigma_{r+1})$$
$$= \text{diag}(\Sigma_1, \sigma_{r+1})$$

with (A, B, C) partitioned conformably with Σ (i.e. A_{11} is (n-1)x(n-1)).

3) Form $g_s^r(z) + g_u(z)$

$$U = -\frac{(C_2 B_2)}{(B_2 B_2)'}$$

$$\Gamma = (\Sigma_1^2 - \sigma_{r+1}^2 I)$$

$$\hat{A} = \Gamma^{-1}(\sigma_{r+1}^2 A_{11}^* + \Sigma_1 A_{11} \Sigma_1 - \sigma_{r+1} C_1^* U B_1^*)$$

$$\hat{B} = \Gamma^{-1}(\Sigma_1 B_1 + \sigma_{r+1} C_1^* U)$$

$$\hat{C} = C_1 \Sigma_1 + \sigma_{r+1} U B_1^*$$

$$\hat{D} = D - \sigma_{r+1} U$$

4) Block diagonalize $\hat{A}$

a) Reduce $\hat{A}$ to real upper Schur form, i.e. find V_1 such that $V_1' V_1 = I$ and $V_1' \hat{A} V_1$ is an upper Schur form.

b) Find an orthogonal matrix V_2 such that

$$V_2' V_1' \hat{A} V_1 V_2 = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ 0 & \hat{A}_{22} \end{bmatrix}$$

where $\mathrm{Re}(\lambda_i(\hat{A}_{11})) < 0$, $\mathrm{Re}(\lambda_i(\hat{A}_{22})) > 0$, and $\hat{A}_{11}$ is r x r.

c) Find $X \in \Re^{r \times (n-r-1)}$ such that

$$\hat{A}_{11} X - X \hat{A}_{22} + \hat{A}_{12} = 0$$

by the Bartels-Stewart algorithm.

d) Let

$$T = V_1 V_2 \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} = (T_1, T_2)$$

$$S = \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} V_2' V_1' = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$$

e) Let

$$\hat{B}_1 = S_1 \hat{B}$$
$$\hat{B}_2 = S_2 \hat{B}$$

$$\hat{C}_1 = \hat{C}\,T_1$$
$$\hat{C}_2 = \hat{C}\,T_2$$

5) Calculation of the D-matrix (term)

a) Find a balanced realization of the system $(-\hat{A}_{22}, \hat{B}_2, \hat{C}_2, \hat{D})$ say $(A_3, B_3, C_3, \hat{D})$ with Hankel singular values $\mu_1 > \mu_2 .... > \mu_{n-r-1} \geq 0$.

b) Let $Z, Y \in \Re^{q\times(n-r-1)}$ be given by

$$Z = \begin{bmatrix} B_3' \\ 0 \end{bmatrix}, \quad Y = \begin{bmatrix} C_3 \\ 0 \end{bmatrix}$$

where q is an integer greater than or equal to (p+m) and let z_i and y_i be the i-th columns of Z and Y respectively.

c) For i=1 to n-r-1

i) Find Householder transformations such that

$$(I - \pi_1^{-1} W_1 W_1')\, y_i = -(\alpha, 0, 0, ..., 0)'$$
$$(I - \pi_2^{-1} W_2 W_2')\, z_i = -(\beta, 0, 0, ..., 0)'$$

ii) Let

$$U := (I - \pi_1^{-1} W_1 W_1') \begin{bmatrix} \frac{-\alpha}{\beta} & 0 & 0 & 0 \\ 0 & 0 & I_{p-1} & 0 \\ 0 & I_{m-1} & 0 & 0 \\ 0 & 0 & 0 & I_{q-p-m+1} \end{bmatrix} (I - \pi_2^{-1} W_2 W_2')$$

iii) If $i < n- r-1$ then for $j := (i+1)$ to $(n-r-1)$

$$y := -(y_j\mu_j + Uz_j\mu_i)(\mu_i^2 - \mu_j^2)^{-\frac{1}{2}}$$

$$z_j := (z_j\mu_j + U'y_j\mu_i)(\mu_i^2 - \mu_j^2)^{-\frac{1}{2}}$$

$$y_j := y$$

iv)

$$\hat{D} := \hat{D} + (-1)^i \mu_i [I_p \quad 0] U \begin{bmatrix} I_m \\ 0 \end{bmatrix}$$

6)

$$g_s^r(z) = \hat{D} + \hat{C}_1 (zI - \hat{A}_{11})^{-1} \hat{B}_1$$

and

$$\|f(z) - g_s^r(z)\|_\infty \le \sigma_{r+1} + \mu_1 + \mu_2 + ... + \mu_{n-r-1}. \qquad (3.81)$$

The above algorithm is designed to handle multi-input, multi-output systems with p representing the number of inputs and m representing the number of outputs. Digital filters are special case where p =1 and m =1.

Note that the singular values defined in step 5,a is used instead of the singular values of H in eq.(3.81). It can be shown that the error bound given by eq.(3.81) is valid and it is less than or equal to the error bound given by eq.(3.78) [33]. Moreover, μ_i represents the i-th singular value of the function $g_u(z^{-1})$ and since $g_u(z)$ is thrown to obtain a stable reduced model, it is logical for μ_i to appear in eq.(3.81).

3.7 General Remarks

In this section, we will discuss the following remarks regarding the above state-space methods for model reduction.

1) In all of the state-space methods discussed above, the approximation error is directly related to the order of the reduced model. The error bound could be calculated beforehand from the singular values of the Hankel matrix depending on the order of the reduced model as given be eq.(3.68). This fact gives the designer the flexibility to select the most suitable order for the reduced model according to the error tolerance of the design. Any simulation program of these methods should include an option for choosing the order of the reduced model after displaying the singular values of the Hankel matrix H.

2) There is no control over the number of poles and zeros of the IIR filter designed using these methods. The numerator degree M and the denominator degree N of the filter are in general equal to the order of approximation r except in the OPH method where M is restricted to be less than or equal to N-1.

3) To have a precise idea about the optimality of optimal Hankel techniques over MS method in time domain, consider the following error bound in Hankel norm [33]:

$$\left\| H(z) - H_r(z) \right\|_H \leq 2 \sum_{i=r+1}^{n} \sigma_i. \qquad (3.82)$$

where $H_r(z)$ is defined by eq.(3.70). The above error bound represents the Hankel norm error of MS method. It indicates two important things. First, the

Hankel norm error bound is equal to the frequency response error bound for MS method. Second, it shows clearly the difference between the MS method and optimal Hankel methods where the error is guaranteed to be exactly equal to σ_{r+1} in Hankel norm.

CHAPTER IV

TWO-SIDED RATIONAL APPROXIMATION METHOD FOR DIGITAL FILTER APPROXIMATION

In this chapter, two-sided method for noncausal digital filters approximation is introduced. The resulting approximation is a stable and causal IIR filter whose magnitude response is extremely close to the magnitude response of the desired filter. We will propose an algorithm to approximate causal digital filters using two-sided method at the end of the chapter. It is a general method where many time domain IIR digital filter design methods could be applied as will be shown later.

4.1 Introduction

All of the well known time domain and frequency domain techniques developed for digital filter design deal with causal digital filters. In the case of time domain techniques, the Fourier coefficients (impulse response) which are obtained from the ideal filter characteristics $H(e^{j\omega})$ are truncated and a constant delay is introduced to obtain a causal impulse response [6]. This limitation is removed in this method. The impulse response is not shifted and the

anticausal part of the impulse response is considered in the approximation process, hence the name two-sided.

This method was originally proposed to approximate the amplitude response of a noncausal digital filter regardless of its phase by a causal and stable IIR filter using the Fourier expansion of the amplitude response [6].

A noncausal filter is a filter whose impulse response h(n) contains some terms with $n < 0$. Theoretically, this means that the present output depends on the past and future inputs. Physically, a noncausal system is unrealizable since in real-time signal processing applications we can't observe future values of the signal. On the other hand, if the signal is recorded so that the processing is done off-line (nonreal-time), it is possible to implement a noncausal filter since all values of the signal are available at the time of processing. This is often the case in the processing of geophysical signals and images [11].

Noncausal digital filters have some desirable features. They have a better computation efficiency compared with linear phase digital filters. Also, they can easily achieve ideal zero phase characteristics [50].

Two-sided approximation technique was first applied using Pade approximation and a formula was obtained for that [6]. In this work, we will apply two-sided approximation to digital filters design using its impulse response as an input. Actually, the impulse response represents the Fourier coefficients of the frequency response expansion. Time domain methods for approximating IIR filters which where discussed in the previous two chapters will be applied to two-sided approximation method and a general formula is derived for any

time domain technique which is applicable to this method. Moreover, a small modification is made so that the two-sided approximation method could handle digital filters with antisymmetric impulse response and it is proved that this modification will not affect the error bound. The efficiency of these methods applied to two-sided approximation will be evaluated in a later chapter.

4.2 Theoretical Basis of Two-Sided Approximation

let $H(e^{j\omega})$ represents the frequency response of an ideal digital filter and let its Fourier expansion be:

$$H(e^{j\omega}) = \sum_{-\infty}^{\infty} h(n)e^{-j\omega n} = h(0) + \sum_{n=1}^{\infty} h(n)e^{-j\omega n} + \sum_{-\infty}^{-1} h(n)e^{-j\omega n} \tag{4.1}$$

where h(n) is assumed to be real and symmetric. Let $\hat{H}(e^{j\omega})$ represents the causal part of the Fourier expansion:

$$\hat{H}(e^{j\omega}) = \frac{1}{2}h(0) + \sum_{n=1}^{\infty} h(n)e^{-j\omega n} \tag{4.2}$$

Suppose that

$$\hat{H}_a(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}}{a_0 + a_1 z^{-1} + \ldots + a_n z^{-n}} \tag{4.3}$$

is a rational approximation of $\hat{H}(e^{j\omega})$ with no poles on $|z| = 1$ and the coefficients of B(z) and A(z) are real. The following theorem provides us with a rational approximation of $H(e^{j\omega})$. Actually, the two-sided approximation method is based on it.

Theorem 4.1: [6]

Suppose that $\hat{H}_a(z) = \frac{B(z)}{A(z)}$ is a rational approximation of $\hat{H}(e^{j\omega})$. Let

$$H_a(z) = \frac{B(z)\,A^*(z)+B^*(z)\,A(z)\,z^{m-n}}{A^2(z)} \tag{4.4a}$$

if $m \leq n$ and let

$$H_a(z) = \frac{B(z)\,A^*(z)\,z^{n-m}+B^*(z)\,A(z)}{A^2(z)} \tag{4.4b}$$

if $m > n$. Then

$$\left|\,|H_a(e^{j\omega})| - |H(e^{j\omega})|\,\right| \leq 2\left|\hat{H}_a(e^{j\omega}) - \hat{H}(e^{j\omega})\right| \tag{4.5}$$

for all ω. Moreover, (4.4a,b) represents a stable filter whenever (4.3) is stable.

m and n represent the orders of B(z) and A(z) respectively. $B^*(z)$ represents the reciprocal polynomial with respect to B(z) and it is given by:

$$B^*(z) = B(z^{-1})\,z^{-m} = b_m + b_{m-1}z^{-1} + \ldots + b_0 z^{-m} \tag{4.6}$$

The same applies for A(z). Before giving the proof of the above theorem, the following lemma is required.

Lemma 4.1: [6]

Let $s^*(z) = a_0 z^{-n} + a_1 z^{-n+1} + \ldots + a_n = a_0 \prod_{i=1}^{n} (z^{-1} - z_i^{-1})$ be the reciprocal polynomial with respect to $s(z)=a_0 +\ldots +a_n z^n$ with real coefficients. Then,

$$s^*(z)=M(z)s(z)$$

where

$$M(z^{-1}) = \prod_{i=1}^{n} \frac{z^{-1}-z_i^{-1}}{1-z^{-1}\,\bar{z}_i^{-1}} \quad . \tag{4.7}$$

proof:

$$s^*(z) = a_0 \prod_{i=1}^{n} (z^{-1} - z_i^{-1}) = a_0 M(z) \prod_{i=1}^{n} (1 - z^{-1}\,\overline{z_i^{-1}}) \tag{4.8}$$

$$= M(z) a_0 \prod_{i=1}^{n} (-\overline{z_i^{-1}}) \prod_{i=1}^{n} (z^{-1} - \bar{z}_i) \tag{4.9}$$

$$= M(z)\,\overline{a_0 \prod_{i=1}^{n} (-z_i^{-1})} \prod_{i=1}^{n} (z^{-1} - \bar{z}_i) \tag{4.10}$$

if z^{-1} is set to zero in the reciprocal polynomial given in lemma 4.1 above, a_n could be written as follows:

$$a_n = a_0 \prod_{i=1}^{n} (-z_i^{-1}) \, . \tag{4.11}$$

Substituting the value of a_n in (4.10)

$$= M(z)\,\overline{a_n} \prod_{i=1}^{n} (z^{-1} - \bar{z}_i) \tag{4.12}$$

$$= M(z)\,\overline{a_n \prod_{i=1}^{n} (\overline{z^{-1}} - z_i)} \tag{4.13}$$

$$= M(z)\,\overline{s(\bar{z})} = M(z)s(z) \tag{4.14}$$

Having the above lemma, we are ready to give the proof of theorem (4.1).

proof: [6]

Using the above lemma, we could write

$$A^*(z) = M(z)A(z).$$

Then, from eq.(4.4)

$$|H_a(z)| = \left|\frac{B(z)A^*(z)+B^*(z)A(z)z^{m-n}}{A(z)A^*(z)M(z)}\right| \tag{4.15}$$

M(z) is actually an all-pass filter and $|M(z)| = 1$ for all $z = e^{j\omega}$. Therefore, for $z = e^{j\omega}$:

$$|H_a(e^{j\omega})| = \left|\frac{B(z)A^*(z)z^n+B^*(z)z^m A(z)}{A(z)A^*(z)z^n}\right| \tag{4.16}$$

$$= \left|\frac{B(z)A(z^{-1})+B(z^{-1})A(z)}{A(z)A(z^{-1})}\right| \tag{4.17}$$

$$= \left|\frac{B(z)\,\overline{A(z)}+\overline{B(z)}\,A(z)}{A(z)\,\overline{A(z)}}\right| \tag{4.18}$$

$$= \left|2\,RE\left[\frac{B(z)}{A(z)}\right]\right| = \left|2\,RE\ \hat{H}_a(e^{j\omega})\right| \tag{4.19}$$

where "RE" stands for the real part of. The above result gives the key to prove the inequality given in eq.(4.5) which is the main result of the theorem. It is proved as follows:

$$\left||H_a(e^{j\omega})| - |H(e^{j\omega})|\right| = \left|\left|2\,RE\ \hat{H}_a(e^{j\omega})\right| - |H(e^{j\omega})|\right| \tag{4.20}$$

$$= \left| \left| 2\,\mathrm{RE}\ \hat{H}_a(e^{j\omega}) \right| - \left| 2\,\mathrm{RE}\ \hat{H}(e^{j\omega}) \right| \right| \tag{4.21}$$

$$\leq \left| \mathrm{RE}(2\ \hat{H}_a(e^{j\omega}) - 2\ \hat{H}(e^{j\omega})) \right| \tag{4.22}$$

$$< 2 \left| \hat{H}_a(e^{j\omega}) - \hat{H}(e^{j\omega}) \right| . \tag{4.23}$$

The importance of the above theorem is summarized in two points: firstly, it gives an explicit expression to approximate $H(e^{j\omega})$ and secondly, it provides the designer with an error bound for the magnitude response. An important aspect to note about this theorem is that any time domain method that provides us with a rational approximation for a desired digital filter could be applied to this theorem. Note that nothing is mentioned about the phase response which is actually affected using this approximation. This could be seen by considering the derivation of eq.(4.4) which is based on the stability and causality requirements of the designed filter. Also, this derivation provides us with a physical interpretation for $H_a(z)$.

4.2.1 Derivation of $H_a(z)$:

Starting from equation (4.1), H(z) could be approximated as follows:

$$H'_a(z) = \hat{H}_a(z) + \hat{H}_a(z^{-1}) \tag{4.24}$$

$$= \frac{B(z)}{A(z)} + \frac{B(z^{-1})}{A(z^{-1})} \tag{4.25}$$

$$= \frac{B(z)A(z^{-1})+A(z)B(z^{-1})}{A(z)A(z^{-1})} \qquad (4.26)$$

where $\hat{H}_a(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})}$ is the approximation of the anticausal part of the impulse response. Note that the anticausal part approximation has the same formulation as the causal part with z replaced by z^{-1}. This is easily justified by the assumption that h(n)= h(-n). Now, if A(z) is stable, then $A(z^{-1})$ is not stable and hence $H_a'(z)$. Therefore, it should be stabilized before using it in practice. Since $|A(z)| = |A(z^{-1})|$ at $z = e^{j\omega}$, $H_a'(z)$ can be stabilized as follows: [45]

$$H_a''(z) = \frac{B(z)A(z^{-1})+B(z^{-1})A(z)}{A(z)A(z)} \qquad (4.27)$$

with $|H_a'(z)| = |H_a''(z)|$.

The numerator of $H_a''(z)$ contains the noncausal terms given by $b_0 a_n z^n$ and $a_0 b_m z^m$. On the other hand, its denominator doesn't contain any term with a positive z [45]. Thus, $H_a''(z)$ is a noncausal function. This can be corrected by introducing the following delay:

$$H_a(z) = \frac{B(z)A(z^{-1})+B(z^{-1})A(z)}{A^2(z)} z^{-n} \qquad \text{if } m \le n \qquad (4.28a)$$

$$= \frac{B(z)A(z^{-1})+B(z^{-1})A(z)}{A^2(z)} z^{-m} \qquad \text{if } m > n\,. \qquad (4.28b)$$

Actually, eq.(4.4a,b) and eq.(4.28a,b) are identical. This can be easily seen by performing simple substitutions using eq.(4.6). Starting from eq.(4.28a), eq.(4.4a) is obtained as follows:

$$H_a(z) = \frac{B(z)A(z^{-1})z^{-n} + B(z^{-1})A(z)z^{-n}}{A^2(z)} \tag{4.29}$$

$$= \frac{B(z)A^*(z) + B^*(z)A(z)z^{m-n}}{A^2(z)} \tag{4.30}$$

where eq.(4.6) is used to substitute $A(z^{-1})z^{-n}$ by $A^*(z)$ and $B(z^{-1})$ by $B^*(z)z^m$ The same argument follows for eq.(4.28b) and eq.(4.4b). Hence, the function $H_a(z)$ has such expression as a consequence of stability and causality.

An important fact to note is that $|H_a(z)| = |H_a'(z)| = |H_a''(z)|$ on the unit circle $z = e^{j\omega}$. Thus, the magnitude is not affected by the causality and stability constrains. However, this is not the case for the phase since poles are added at $z = 0$ to achieve causality and half the poles of the denominator are reciprocated to achieve stability. Therefor, this method is not expected to give zero phase characteristic as will be shown later.

4.2.2 Antisymmetric Impulse Response

Noncausal digital filters design with antisymmetric impulse response were first handled in [50] based on frequency domain specifications. The above method could be modified so that it can also handle antisymmetric impulse response where h(n) = -h(-n). This can be simply done by replacing the second term in eq.(4.24) by the function $-\hat{H}_a(z^{-1})$ which represents the frequency response of h(-n), the anticausal part of the impulse response. Consequently, eq.(4.4a,b) is modified as follows:

$$H_a(z) = \frac{B(z)A^*(z) - B^*(z)A(z)z^{m-n}}{A^2(z)} \tag{4.31a}$$

$$H_a(z) = \frac{B(z)A^*(z)z^{n-m} - B^*(z)A(z)}{A^2(z)} \qquad (4.31b)$$

The above modification is necessary since the method is based on the assumption that h(n) = h(-n) which leads to the symmetry of the functions $\hat{H}_a(z)$ and $\hat{H}_a(z^{-1})$. The symmetry of these two functions make the derivation and analysis of this method much easier than other methods. Moreover, it enables us to find the error bound of the magnitude response as described above. In case of antisymmetric impulse response, this symmetry is preserved by utilizing the relation h(n) = -h(-n) where h(n) in the second term of eq.(4.1) is substituted by -h(n) (for clarification, h(n) in the second term of eq.(4.1) is referred her as h(-n)). The error bound given by eq.(4.5) is not affected by this modification. This can be shown by considering the proof of theorem 4.1 starting from eq.(4.18)

$$\left|H_a(e^{j\omega})\right| = \left|\frac{B(z)\,\overline{A(z)} - \overline{B(z)}\,A(z)}{A(z)\,\overline{A(z)}}\right| \qquad (4.32)$$

$$= \left|2\,\mathrm{IM}\left[\frac{B(z)}{A(z)}\right]\right| = \left|2\,\mathrm{IM}\ \hat{H}_a(e^{j\omega})\right| \qquad (4.33)$$

where "IM" stands for the imaginary part of. Thus :

$$\left|\left|H_a(e^{j\omega})\right| - \left|H(e^{j\omega})\right|\right| = \left|\left|2\,\mathrm{IM}\ \hat{H}_a(e^{j\omega})\right| - \left|2\,\mathrm{IM}\ \hat{H}(e^{j\omega})\right|\right| \qquad (4.34)$$

$$\leq \left| IM(2\,\hat{H}_a(e^{j\omega}) - 2\,\hat{H}(e^{j\omega})) \right| \tag{4.35}$$

$$< 2\left| \hat{H}_a(e^{j\omega}) - \hat{H}(e^{j\omega}) \right| . \tag{4.36}$$

4.3 General Formulae for the Coefficients of $H_a(z)$

In this section, we are going to give a general formulae for the coefficients of C(z) and D(z) in terms of B(z) and A(z).

Let $H_a(z)$ be written as:

$$H_a(z) = \frac{C(z)}{D(z)} \tag{4.37}$$

The order of C(z) is n+m+abs(m-n) where abs(m-n) is a shift caused by the causality factor z^{m-n} or (z^{n-m}) and D(z) is of order 2n. We will use the following notation for simplification:

$$c_1(i) = \sum_{j=0}^{i} b_j a_{n-i+j} \qquad b_j = 0 \text{ for } j > m \tag{4.38}$$

$$c_2(i) = \sum_{j=0}^{i} a_j b_{m-i+j} \qquad a_j = 0 \text{ for } j > n \tag{4.39}$$

where b_i's and a_i's are the coefficients of B(z) and A(z) respectively. The coefficients of C(z) and D(z) denoted as c(i) and d(i) respectively are given as follows :

i) Symmetric h(n):

for $m \leq n$:

$$c(i) = c_1(i) \qquad 0 \le i < n-m \qquad (4.40a)$$
$$= c_1(i) + c_2(i-n+m) \qquad n-m \le i \le m+n \qquad (4.40b)$$
$$= c_2(i-n+m) \qquad m+n < i \le m+n+(n-m) \qquad (4.40c)$$

and $$d(i) = \sum_{j=0}^{i} a_j a_{i-j} \qquad 0 \le i \le 2n. \qquad (4.41)$$

for $m > n$ *:*

$$c(i) = c_2(i) \qquad 0 \le i < m-n \qquad (4.42a)$$
$$= c_1(i-m+n) + c_2(i) \qquad m-n \le i \le m+n \qquad (4.42b)$$
$$= c_1(i-m+n) \qquad m+n < i \le m+n+(m-n) \qquad (4.42c)$$

and d(i) is given by eq.(4.41).

ii) Antisymmetric h(n):

for $m \le n$ *:*

$$c(i) = c_1(i) \qquad 0 \le i < n-m \qquad (4.43a)$$
$$= c_1(i) - c_2(i-n+m) \qquad n-m \le i \le n+m \qquad (4.43b)$$
$$= -c_2(i-n+m) \qquad n+m < i \le n+m+(n-m) \qquad (4.43c)$$

for $m > n$ *:*

$$c(i) = -c_2(i) \qquad 0 \le i < m-n \qquad (4.44a)$$
$$= c_1(i-m+n) - c_2(i) \qquad m-n \le i \le m+n \qquad (4.44b)$$
$$= c_1(i-m+n) \qquad m+n < i \le m+n+(m-n) \qquad (4.44c)$$

for d(i), it is the same as given in eq.(4.41).

The derivation of the above formulae is based on the observation that $B(z)A^*(z^{-1})$ and $B^*(z^{-1})A(z)$ in eq.(4.4a,b) have the same coefficients but one is the reciprocal of the other and the effect of the term z^{m-n} (or z^{n-m}) is merely a shift to the left.

The most important thing about the above formulae is its generality. It can be applied to any method which is suitable for two-sided approximation.

4.4 Algorithm

In this section, we are going to propose a simple algorithm to utilize two-sided method in the approximation of a causal IIR or FIR digital filter by a causal and stable IIR digital filter. It should be noted that two-sided approximation developed above is applicable for digital filters whose impulse response is either symmetric or antisymmetric with odd length. In this algorithm, it is assumed that the impulse response of the causal digital filter is given and it satisfies the conditions stated above. The steps of the algorithm are as follows:

1) Shift the causal impulse response h(n) to the left by (k-1)/2 where k is the length of the impulse response. Denote the shifted version (noncausal) of h(n) by h'(n).

2) Apply eq.(4.3) to $h'_c(n)$, the causal part of h'(n) to obtain the function $\hat{H}_a(z)$. Note that $h_0/2$ should be used instead of h_0 as indicated by eq.(4.2). Any time domain method that gives a rational approximation could be applied in eq.(4.3) to approximate $\hat{H}(z)$.

For h(n) symmetric:

3) Apply eq.'s (4.40) or (4.42), depends on the value of m and n, and eq.(4.41) to obtain the two-sided approximation model $H_a(z)$.

For h(n) antisymmetric:

3') Apply eq.'s (4.43) or (4.44), depends on the value of m and n, and eq.(4.41) to obtain the two-sided approximation on model $H_a(z)$.

Note that step 1 is not necessary if the given impulse response is already noncausal.

4.5 REMARKS

The following remarks should be noted about two-sided approximation:

1) It might be thought that the model $H_a(z)$ obtained using this method is of a very high order ($H_a(z)$ is of order (m+n)) which means more complexity and cost. However, this is not necessary true since it shows a very excellent performance (for magnitude response) with a very low order values of m and n compared with other methods as will be shown later.

2) The phase response is affected because of: 1) the shift of the impulse response to obtain the noncausal version h'(n), 2) the stability and causality requirements of $H_a(z)$. Thus neither linear phase nor zero phase characteristic could be achieved using this method. Two-sided approximation is best applied when phase response is not of much importance to the designer.

3) This method is limited since it can deal only with digital filters whose impulse response has a specific features.

4) Two-sided approximation technique will save much CPU time since it deals with half of h(n). Other techniques usually take the whole impulse response for the approximation process. This means that matrices with high dimension are formed which consume large CPU time for their manipulation.

CHAPTER V

SIMULATION AND EXAMPLES

5.1 Introduction

This chapter is totally devoted for simulation and examples of the previously discussed methods. Prony method and OPHD method are originally available as M-files in Matlab package. However, OPHD method requires some steps so that it can be directly applied to digital filter design. The following methods are implemented and provided as interactive M-files which can be utilized by any user: Pade method, Shank method, Kung method, Kimura method, MS method, CF method, OPH method, OPHd method, OPHD method, and two-sided approximation technique. A copy of the M-files are provided at appendix A.

The chapter is divided into two main parts. In the first part, we will provide several examples to evaluate and compare the performance of the above methods in approximating and designing IIR digital filters. Also, the approximation of LPH FIR filters will be examined closely. Moreover, advantages and disadvantages of each method is highlighted during the discussion. In the second part, the efficiency of two-sided approximation technique in upgrading the performance of the previous methods will be shown through different examples.

The methods are classified into three groups to reduce the number of graphs required for each example. These groups are : least squares methods, suboptimal methods which include suboptimal Hankel methods with MS method, and optimal Hankel methods. Thus, the frequency response and impulse response obtained from each group will be shown in one graph only. The response of the desired or approximated filter will be shown in a solid line in all graphs.

5.2 Comparative Study

In this section, six examples will be introduced for comparison between the above methods and general conclusions will be given at chapter 6.

5.2.1 Example 1: NLPH LPF [13,28]

Consider the following transfer function of a (4,4) IIR filter which will be approximated by a 4-th order (r = 4) and a 2-nd order (r = 2) IIR filters:

$$H(z) = \frac{1+4z+6z^2+4z^3+z^4}{1-1.25398z+0.98713z^2-0.34093z^3+0.05237z^4}$$

The first 20 samples of the impulse response of the above filter given in table 1 are used as input.

The performance of least squares methods: Pade, Prony, and Shank is shown in Fig. 5.1(a,b,c) for r = 2 and in Fig. 5.2(a,b,c) for r = 4. Actually,

they give an excellent approximation for both frequency response H(z) and impulse response h(n) when r = 4. This is logical since the order of the designed filter matches the order of the approximated filter. However, the performance is poor when r = 2 specially for Pade approximation. Fig. 5.1a shows the main disadvantage of Pade where it exactly matches the first 5 samples of h(n) but it deviates greatly for n > 4. This is obvious from the LSE given in table 2.

For suboptimal methods, there many important things to note. Firstly, the impulse response of the IIR filters designed using Kung method starts with 0 at n = 0 for both r = 4 and r = 2 as shown in Fig. 5.3a and 5.4a. This is due to the D-term which is neglected in this method. On the other hand, the impulse response obtained by Kimura method starts with the correct value at n = 0 since the D-term is forced to be equal to h(0). Secondly, the MS method gives a much better approximation than Kung and Kimura methods for the magnitude response at low frequencies (pass-band) when r = 2 as shown in Fig. 5.3b. However, Kung and Kimura have a better performance at high frequencies (stop-band). Thirdly, Kimura and MS methods perfectly match the magnitude response when r = 4 which is not the case for Kung method due to the D-term effect which is shown in Fig. 5.4b. Finally, it is clear from Fig. 5.3c and Fig. 5.4c that the phase response given by Kimura method has a better linear characteristic compared with the phase responses obtained by Kung and MS methods when r = 4.

The performance of OPH, OPHd, OPHD, and CF methods is shown in Fig. 5.5(a,b,c) and Fig. 5.6(a,b,c). h(n) of these methods is identical for both

orders r = 2 and r = 4 except at t = 0 where the impulse response of the OPH method is equal to 0. This is again because of the D-term which is also neglected in this technique. Note that OPHd method starts with the correct value at n = 0. The magnitude and phase responses of OPHD and CF methods are almost identical for both orders as shown in Fig. 5.5(b,c) and Fig. 5.6(b,c). The situation is different for OPH method. When r = 2, it shows a better performance at the stop-band compared to OPHD and CF methods but this is not the case for the pass-band where the performance of OPHD and CF methods is better as shown in Fig. 5.5b. When r = 4, OPH is less efficient in both bands. Note that the L_∞ error norm given in table 2 for OPH is approximately equal to the DC-term h(0) when r = 4. On the other hand, OPHd has the least L_∞ error of all methods. The magnitude and phase responses of OPHd method is better specially when r = 4 compared to the responses of OPH method except at the stop-band where the magnitude response of OPH method is better when r = 2. Actually, when r = 4 OPHd method gives the best approximation compared to all methods which could be seen from the frequency response error given in table 2.

In general, the above methods give a very efficient IIR filter design when r = 4. However, the performance of suboptimal and optimal Hankel techniques is superior compared to the performance of least squares methods when r = 2. Another important thing to note is that the methods which neglect the D-term, namely: Kung and OPH methods, suffer from a large LSE and L_∞ error norm compared to the error norms of the other methods as given in table 2. It will shown later that this disadvantage is not effective when h(0) is equal or close to 0.

TABLE 1

Impulse Response and Singular Values of Example 1.

Impulse response	Singular values
1.00000000000000	31.16372397244300
5.25398000000000	17.38086842296685
11.60125584040000	4.65702364842015
13.70231152134479	0.44794808598402
8.46934632520189	0.03610043167231
0.77449335387915	0.03401208337596
-3.32517536348883	0.02704900513833
-2.76436483844420	0.02576299302069
-0.36356951146433	0.02124004501783
1.09866631337048	0.02049147556312
0.96828048490723	0.01856541809855
0.15049591759235	0.01723960444953
-0.37349140279123	0.01600532479139
-0.34433107351688	0.01599324829566
-0.06249998694122	0.01516704468208
0.12630990381824	0.01465794548935
0.12225265716935	0.01435426420100
0.02534258945333	0.00123442438779
-0.04256420532403	0.00009233305987
-0.04332634377351	

TABLE 2

LSE and L_∞ Error Norms of all Methods Applied in Example 1.

		r = 2		r = 4	
		LSE	L_∞	LSE	L_∞
Least squares methods	Shank	6.41203101	10.78120788	$0.97890776 \times 10^{-12}$	0.04292365
	Prony	6.47247865	10.57224245	$0.13523271 \times 10^{-12}$	0.04292365
	Pade	2.87502471×10^{2}	$60.35702268 \times 10^{2}$	$0.00224835 \times 10^{-12}$	0.04292365
Suboptimal methods	Kung	4.46476009	6.89400330	1.00000101	1.04001172
	Kimura	4.35133114	6.53425430	0.00142320	0.04052739
	MS	7.06340721	9.00047366	0.01217782	0.05081324
Optimal Hankel methods	OPH	4.69829623	6.38966921	1.00026778	1.04422779
	OPHd	4.59064129	5.42056176	0.02314356	0.03286037
	OPHD	4.65518298	4.78023421	0.02316852	0.04583067
	CF	4.65524377	4.78024460	0.02315887	0.04591585

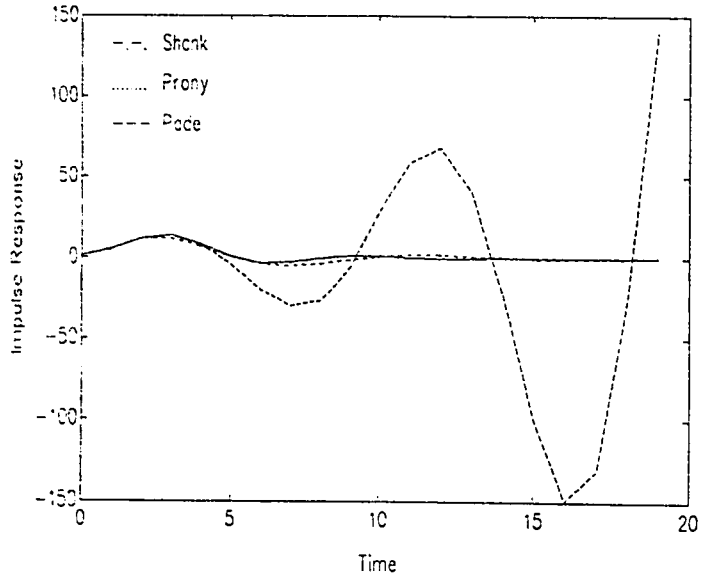


Fig. 5.1a : Impulse response of least squares methods with r = 2.

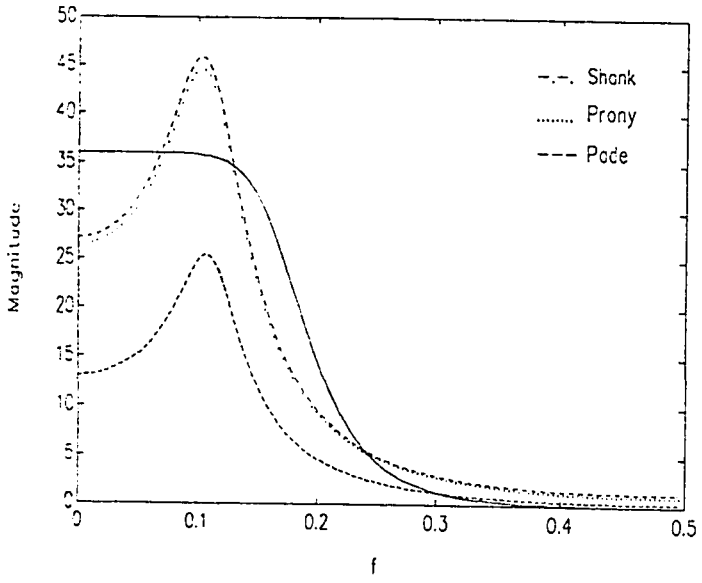


Fig. 5.1b: Magnitude response of least squares methods with r = 2.

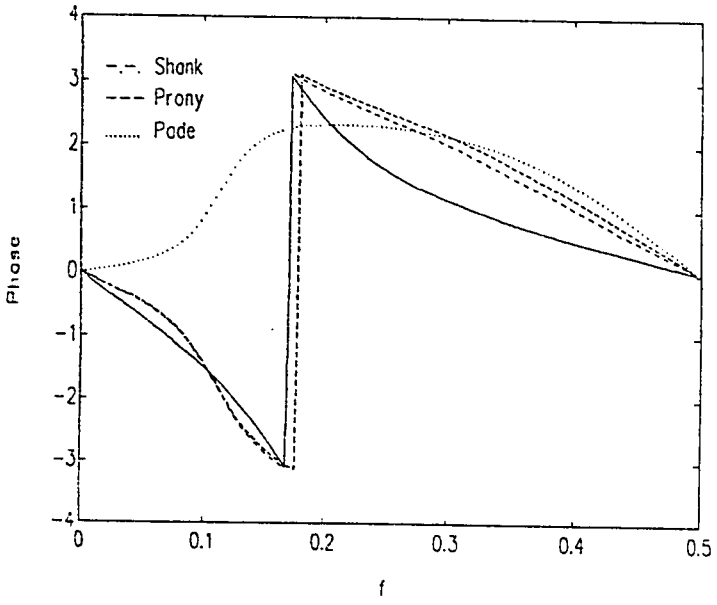


Fig. 5.1c: Phase response of least squares methods with r = 2.

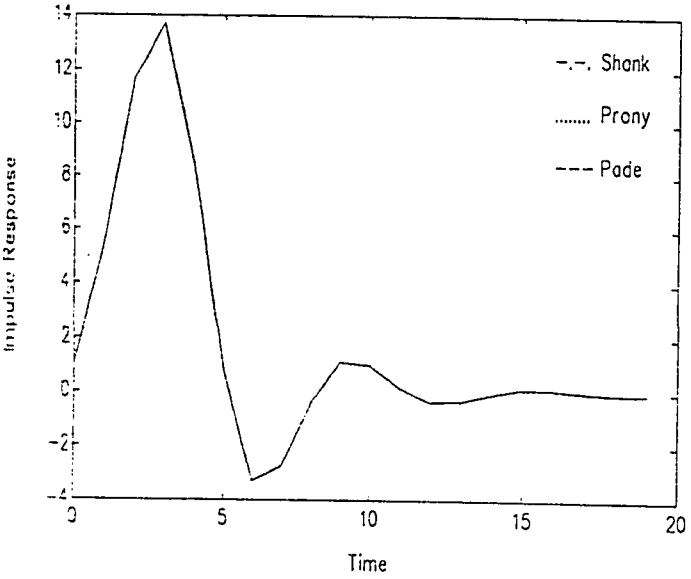


Fig. 5.2a: Impulse response of least squares methods with r = 4.

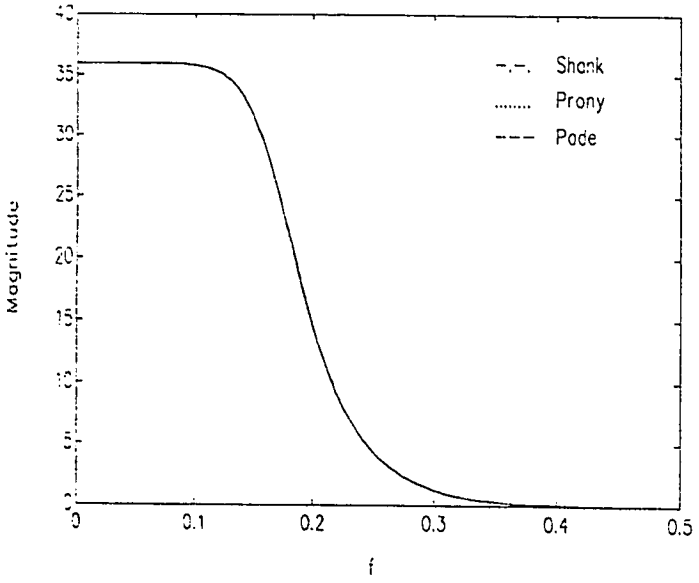


Fig. 5.2b: Magnitude response of least squares methods with r = 4.

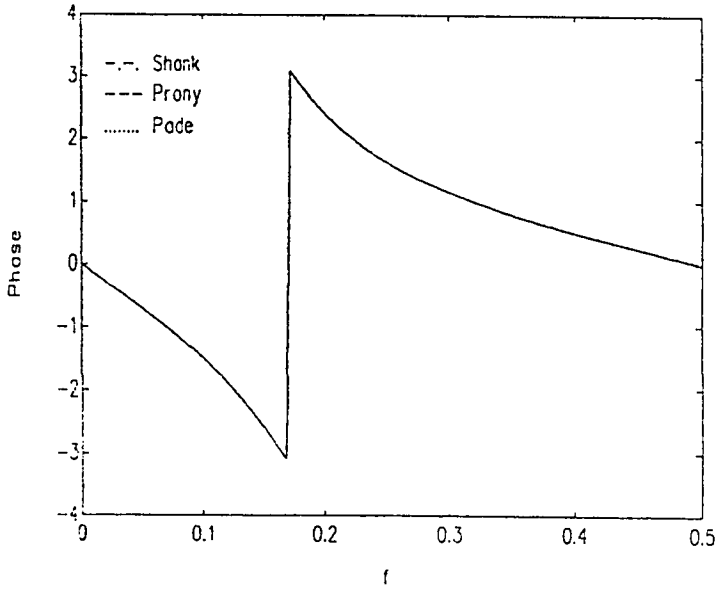


Fig. 5.2c: Phase response of least squares methods with r = 4.

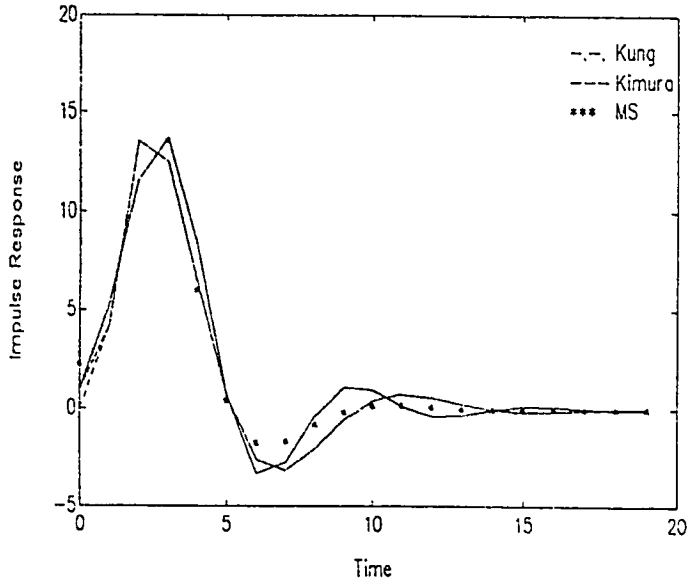


Fig. 5.3a: Impulse response of suboptimal methods with r = 2.

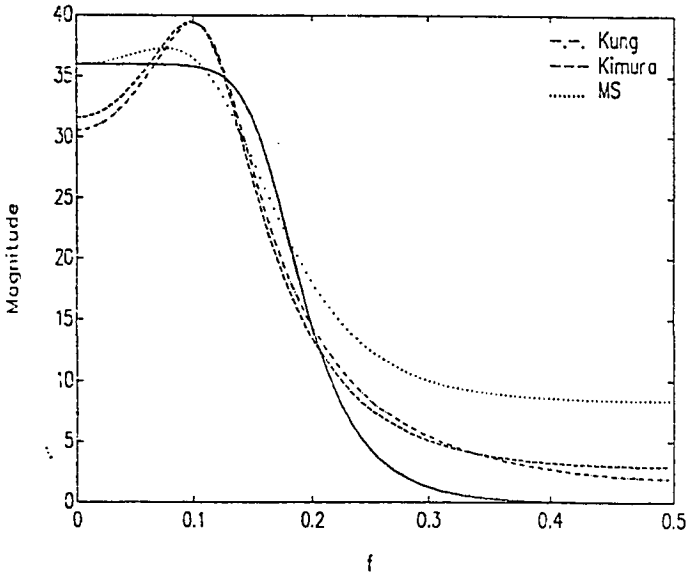


Fig. 5.3b: Magnitude response of suboptimal methods with r = 2.

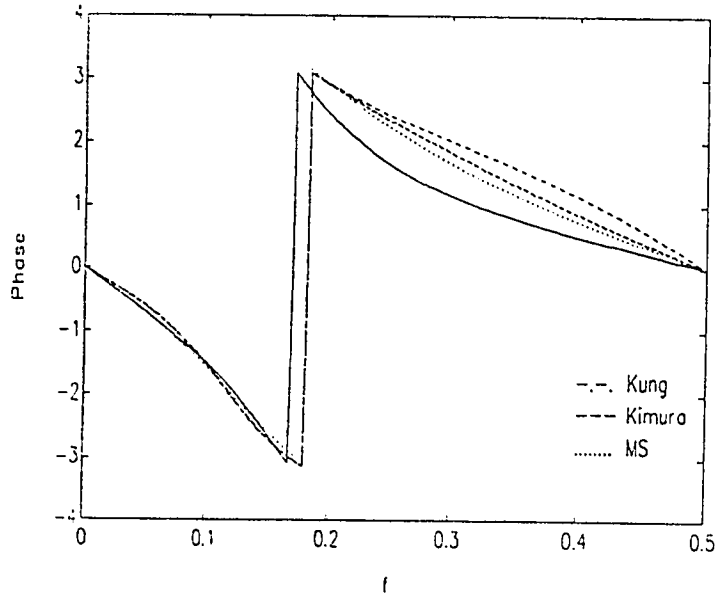


Fig. 5.3c: Phase response of suboptimal methods with r = 2.

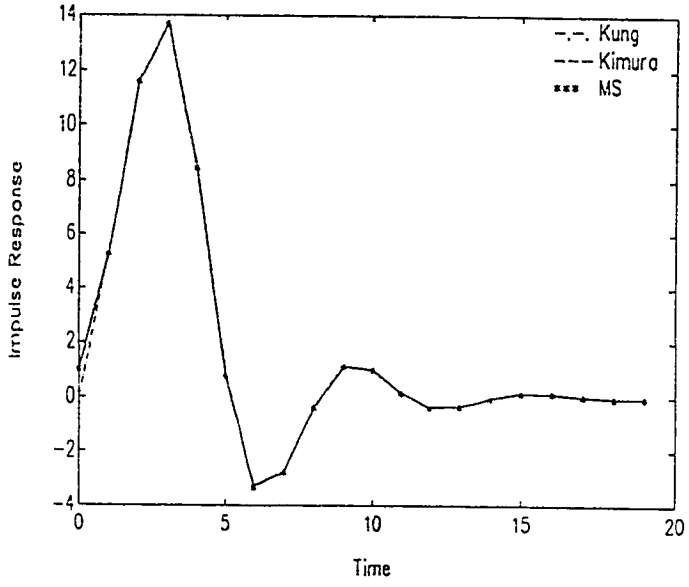


Fig. 5.4a: Impulse response of suboptimal methods with r = 4.

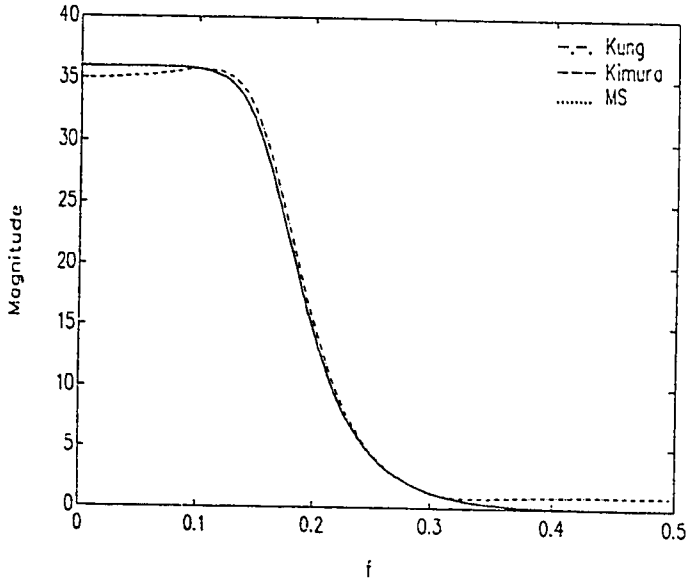


Fig. 5.4b: Magnitude response of suboptimal methods with r = 4.

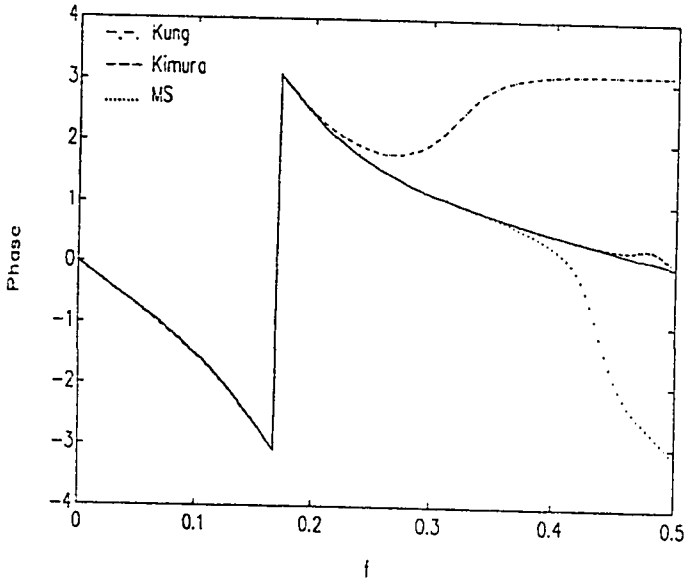


Fig. 5.4c: Phase response of suboptimal methods with r = 4.

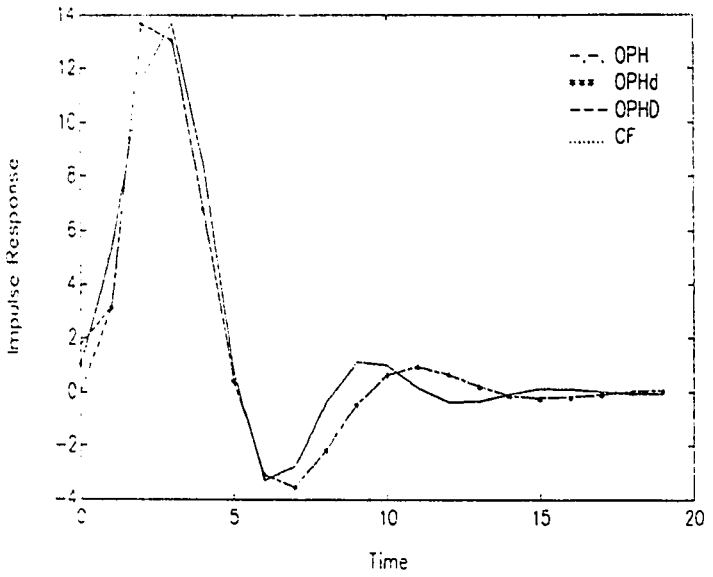


Fig. 5.5a: Impulse response of optimal Hankel methods with r = 2.

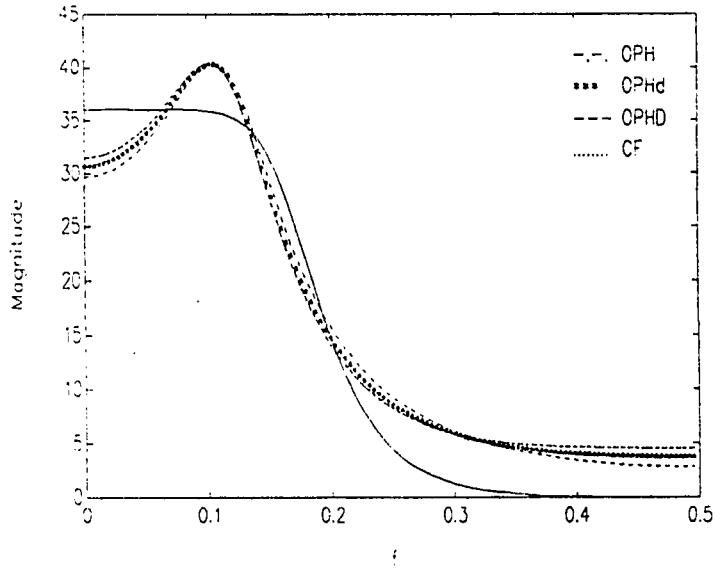


Fig. 5.5b: Magnitude response of optimal Hankel methods with r = 2.

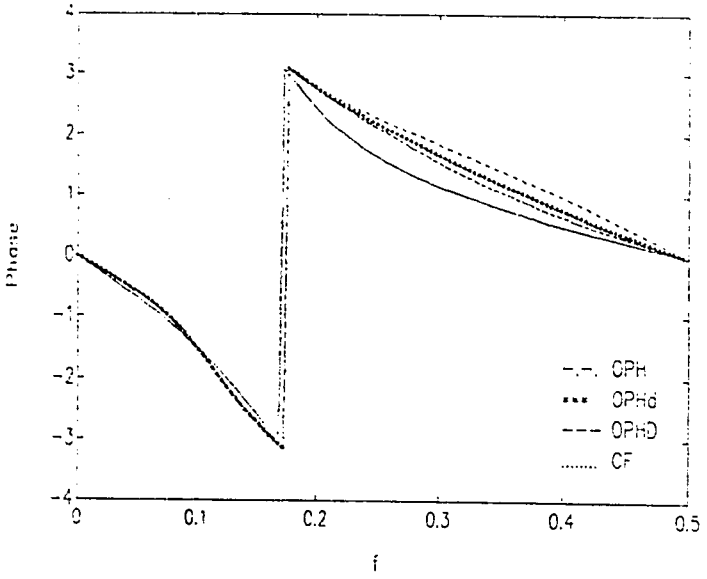


Fig. 5.5c: Phase response of optimal Hankel methods with r = 2.

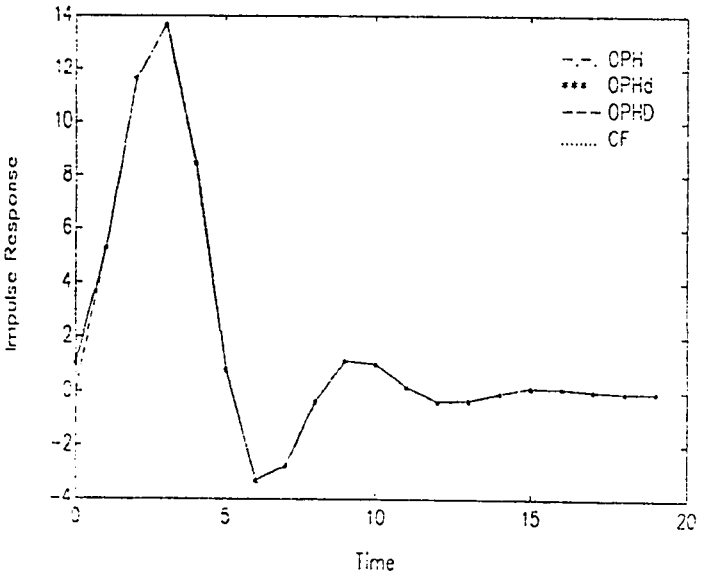


Fig. 5.6a: Impulse response of optimal Hankel methods with r = 4.

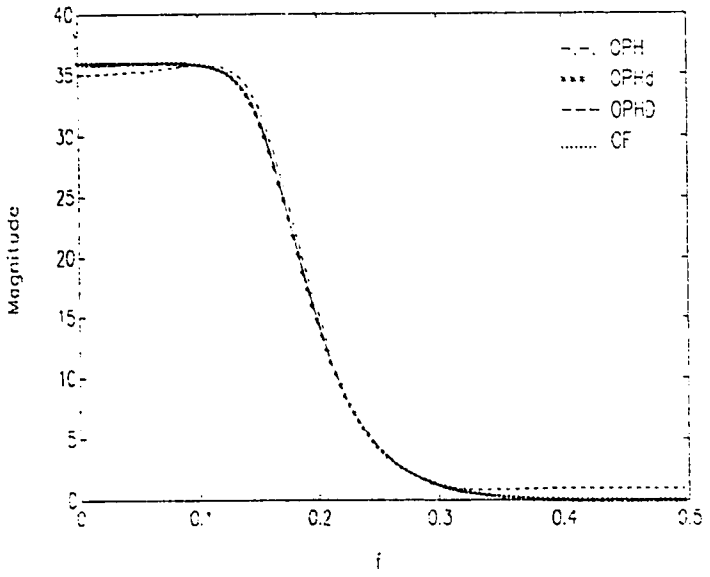


Fig. 5.6b: Magnitude response of optimal Hankel methods with r = 4.

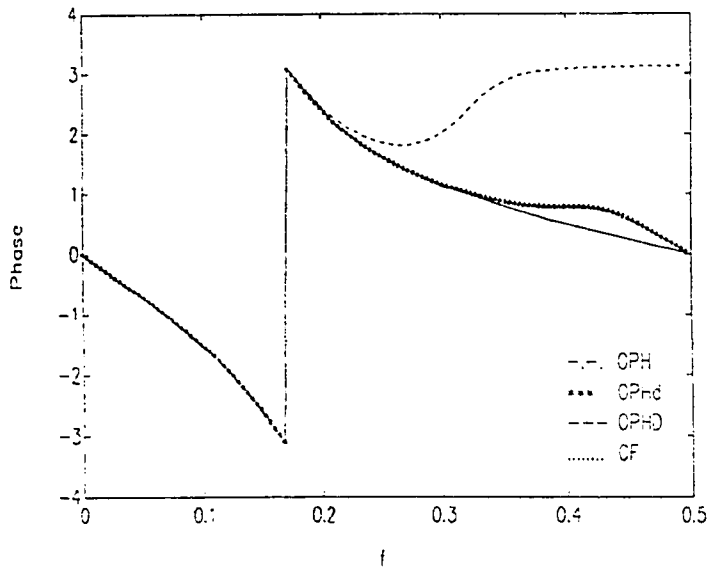


Fig. 5.6c: Phase response of optimal Hankel methods with r = 4.

5.2.2 Example 2: LPH LPF [15]

The impulse response of a LPH FIR filter with $f_c = 0.2$ designed using Remez algorithm is listed in table 3. It is approximated by an (5,5) and (7,7) IIR filters.

Least squares methods give an excellent approximation to the impulse response when r = 7 except for Pade approximation where it starts to deviate increasingly after n = 15. This is shown in Fig. 5.8a. For r = 5, Shank method gives a much better approximation to the impulse response compared to the other two methods as shown in Fig. 5.7a and table 6 where the LSE error is provided. The magnitude response and phase response of the designed IIR filters deviate greatly from the magnitude and phase responses of the original FIR filter as shown in Fig 5.7(b,c,d) and Fig. 5.8(b,c,d). However, the frequency response approximation of the (5,5) IIR filter is better than frequency response approximation of the (7,7) IIR filter. This could be also seen from table 6 where L_∞ error norm is less for r = 5. This seems not to be logical. Actually, the answer is found by considering the poles of the designed IIR filters shown in table 4 and table 5. For r = 7, the designed IIR filters are unstable. On the other hand, Shank and Prony give a stable design for r = 5. This indicates one of the main disadvantages of least squares methods where they may result in an unstable filter.

Reasonable approximations are obtained when suboptimal methods are applied to this example which could be seen from Fig. 5.9(a,b,c,d), Fig. 5.10(a,b,c,d), and table 6. The important observation is the efficiency of the MS method in approximating the magnitude response in the pass-band where

it shows a better performance compared to Kung and Kimura methods. However, the situation is different in the stop-band where Kung and Kimura methods give a better approximation especially for r = 5. The impulse response and magnitude response of Kung and Kimura methods are almost identical for both orders r = 5 and r = 7. Note that h(0) is small in this case. The phase response is shown in Fig. 5.9d for r = 5 where all of the three methods showed a linear behaviour specially at low frequencies. For r = 7, MS method and Kung method showed a better linearity of phase compared to Kimura at high frequencies.

The performance of OPH, OPHd, OPHD, and CF methods are shown in Fig. 5.11(a,b,c,d) and Fig. 5.12(a,b,c,d). OPHd, OPHD method and CF method have an almost identical frequency response and impulse response when r = 5. The impulse response of OPH method differs at n = 0 where h(0) = 0 for r = 5 and r = 7. The magnitude response of the these methods is comparable when r = 5. However, OPH and OPHd methods has more attenuation at the stop-band for both orders. CF method has a less efficient approximation for magnitude response when r = 7 as shown in Fig. 5.12b. The phase response is linear at low frequencies but it deviates slightly at high frequencies when r = 5 as shown in Fig. 5.11d. The deviation from linearity at high frequencies is more apparent with r = 7 except for OPH method which shows a much better phase linearity compared with other three methods.

In the following discussion, the effect of the conversion of the nonparametric impulse response to a parametric form via Prony method on the optimality of the CF method is investigated. A (6,7) and (8,8) IIR filters were

designed for illustration. The results are shown in Fig. 5.13(a,b,c). It seems surprising to find that the (6,7) IIR filter gives a better approximation than the (8,8) IIR filter which is theoretically not expected. This is actually due to the conversion of the impulse response to a parametric form using Prony method which could result in a poor matching to the impulse response. To confirm this analysis, LSE and L_∞ error norm due to this conversion are calculated for both IIR filters and listed in table 7. The error norms for (8,8) IIR filter are considerable but they are negligible for (6,7) IIR filter. Actually, this indicates the a main disadvantage of the CF method since its performance is highly effected by the efficiency of Prony method in matching a certain impulse response.

TABLE 3

Impulse Response and Singular Values of Example 2.

Impulse response	Singular values
- 0.00241346067625	0.99758994144429
0.01100759767901	0.95674335000531
0.01621018603147	0.76805585052153
0.00755098687883	0.43003558804231
- 0.01952976669310	0.17296082900793
- 0.04651412308345	0.05678955797945
- 0.03852398641700	0.01881573472595
0.02921311395140	0.00827210522725
0.14416117167923	0.00610427203329
0.25454562233312	0.00582681810506
0.30019171215131	0.00581172903642
0.25454562233312	0.00581007177791
0.14416117167923	0.00580799308465
0.02921311395140	0.00580707092919
- 0.03852398641700	0.00580497149671
- 0.04651412308345	0.00580383317884
- 0.01952976669310	0.00580336361465
0.00755098687883	0.00007394815858
0.01621018603147	0.00001589826721
0.01100759767901	0.00000000000000
- 0.00241346067625	

TABLE 4

Poles of (5,5) IIR Filter Designed Using Least Squares Methods Applied to the Impulse Response of Example 2.

Shank	Prony	Pade
0.60201703 + 0.74616493i	0.60201703 + 0.74616493i	0.65612604 + 0.98726404i
0.60201703 - 0.74616493i	0.60201703 - 0.74616493i	0.65612604 - 0.98726404i
0.94541734	0.94541734	1.29703290 + 0.65757916i
0.82769565 + 0.46803606i	0.82769565 + 0.46803606i	1.29703290 - 0.65757916i
0.82769565 - 0.46803606i	0.82769565 - 0.46803606i	0.71454641

TABLE 5

Poles of (7,7) IIR Filter Designed Using Least Squares Methods Applied to the Impulse Response of Example 2.

Shank	Prony	Pade
-1.04447072	-1.04447072	0.08234645 + 0.82239024i
0.55631069 + 0.84645406i	0.55631069 + 0.84645406i	0.08234645 - 0.82239024i
0.55631069 - 0.84645406i	0.55631069 - 0.84645406i	1.21938987
0.82016505 + 0.60658863i	0.82016505 + 0.60658863i	0.99408452 + 0.61035214i
0.82016505 - 0.60658863i	0.82016505 - 0.60658863i	0.99408452 - 0.61035214i
1.00217040 + 0.22378093i	1.00217040 + 0.22378093i	0.61497702 + 0.89087619i
1.00217040 - 0.22378093i	1.00217040 - 0.22378093i	0.61497702 - 0.89087619i

TABLE 6

LSE and L_∞ Error Norms of all Methods Applied in Example 2.

		r = 5		r = 7	
		LSE	L_∞	LSE	L_∞
Least squares methods	Shank	0.09186050	0.99166547	0.00069440	3.50893074
	Prony	0.33590728	0.83654142	0.00511617	3.49220598
	Pade	6.98962785	1.04889011	1.42684496	1.16564341
Suboptimal methods	Kung	0.03496671	0.09335050	0.00545898	0.01436927
	Kimura	0.03488333	0.09553979	0.00489650	0.01297930
	MS	0.07637658	0.12428300	0.00917061	0.01759099
Optimal Hankel methods	OPH	0.04461212	0.07336235	0.00528282	0.01252139
	OPHd	0.04454678	0.07113450	0.00469930	0.01011242
	OPHD	0.04637273	0.06562845	0.00683131	0.01474083
	CF	0.04564675	0.06274871	0.01309302	0.04099366

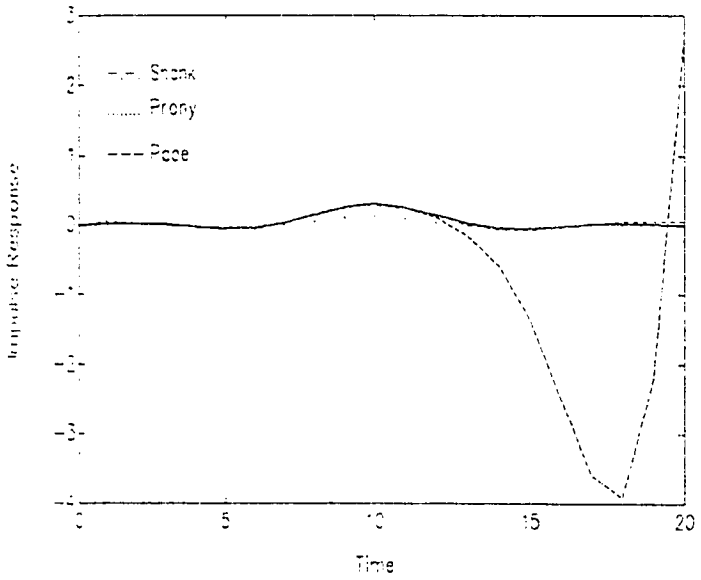


Fig. 5.7a: Impulse response of least squares methods with r = 5.

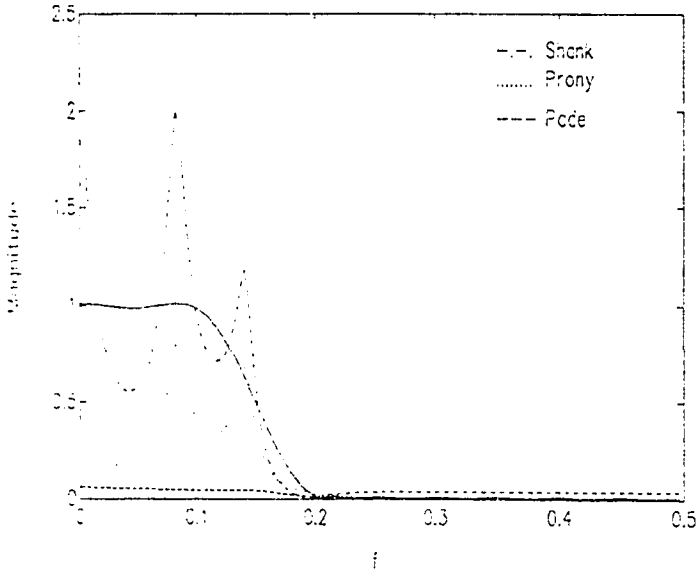


Fig. 5.7b: Magnitude response of least squares methods with r = 5.

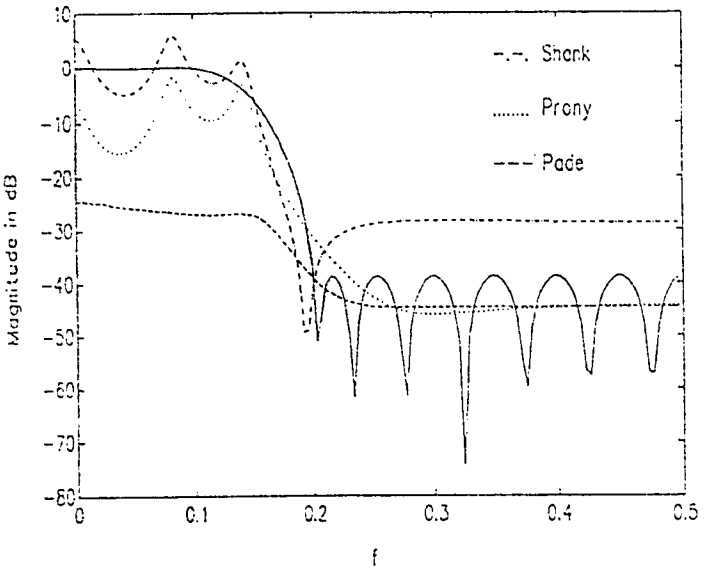


Fig. 5.7c: Magnitude response in dB of least squares methods with r = 5.

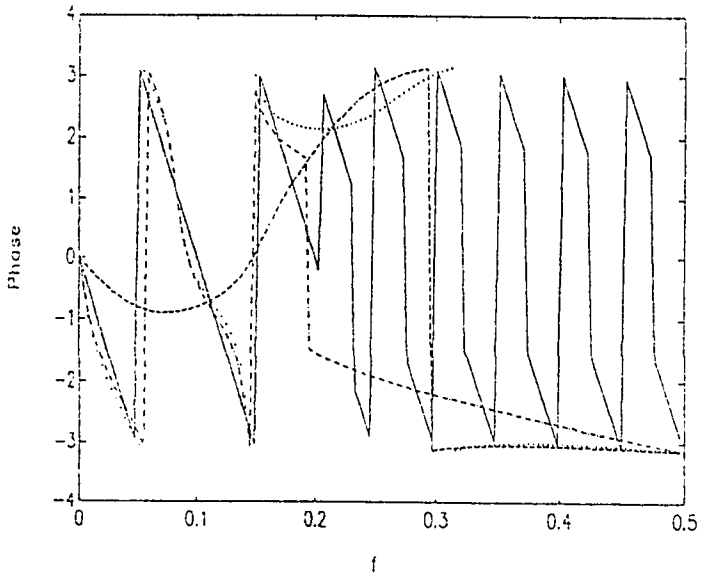


Fig. 5.7d: Phase response of least squares methods with r = 5.

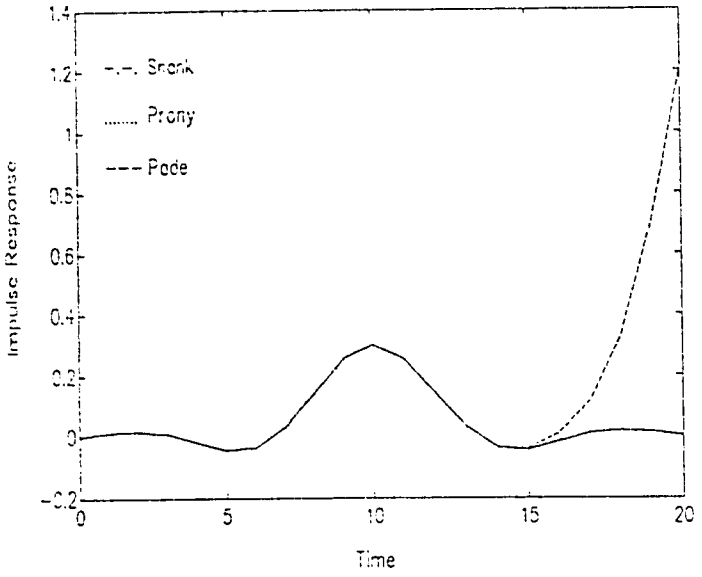


Fig. 5.8a: Impulse response of least squares methods with r = 7.

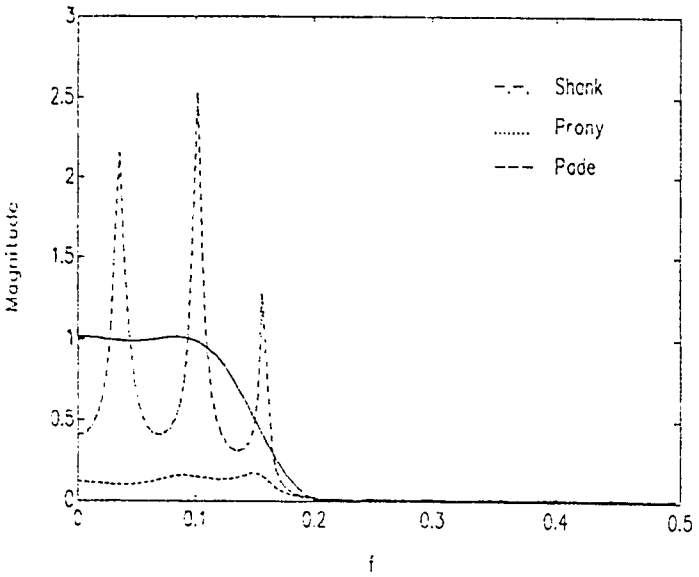


Fig. 5.8b: Magnitude response of least squares methods with r = 7.

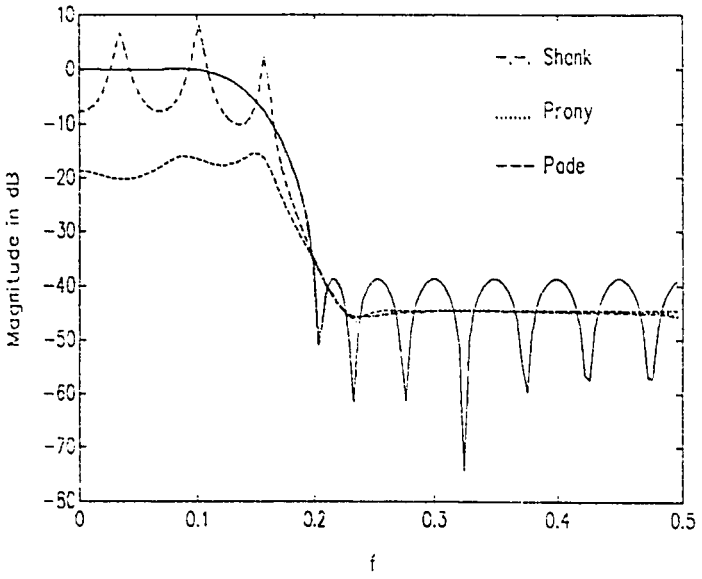


Fig. 5.8c: Magnitude response in dB of least squares methods with r = 7.

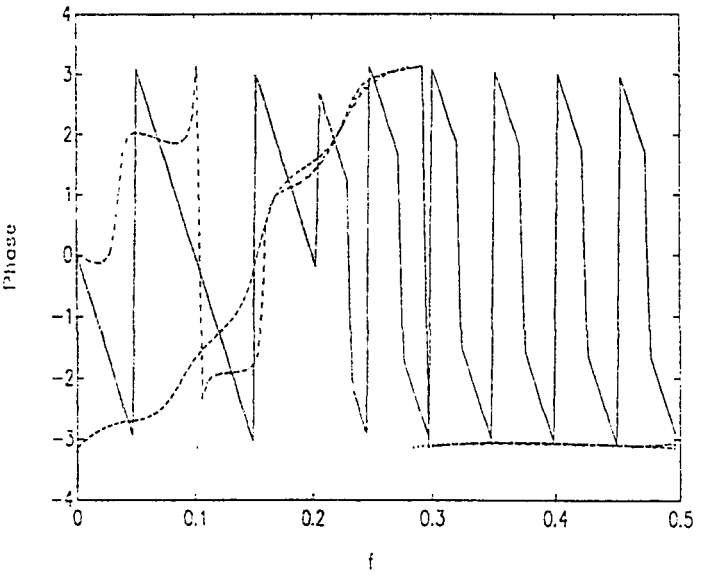


Fig. 5.8d: Phase response of least squares methods with r = 7.

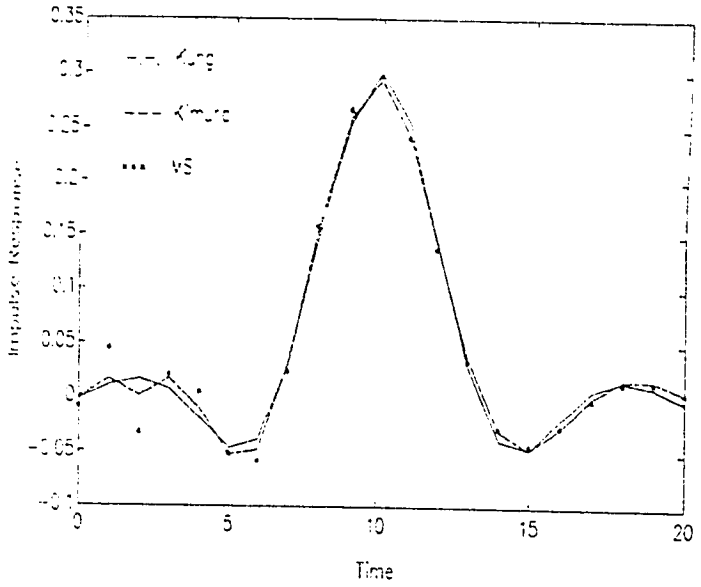


Fig. 5.9a: Impulse response of suboptimal methods with r = 5.

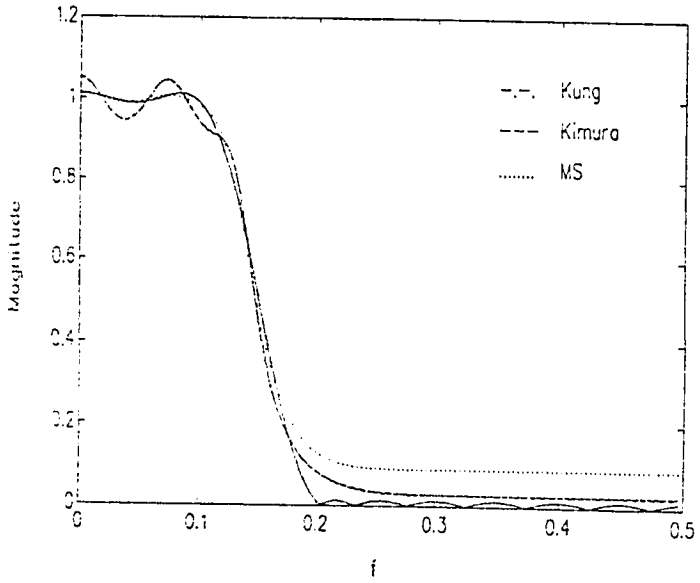


Fig. 5.9b: Magnitude response of suboptimal methods with r = 5.

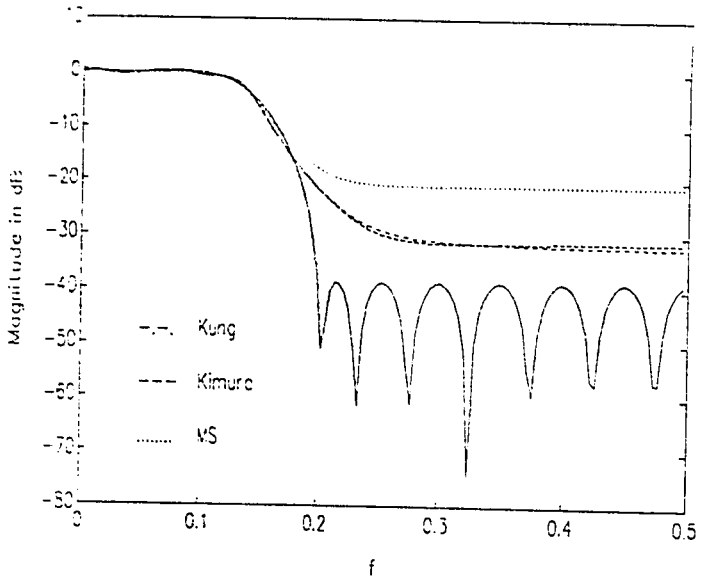


Fig. 5.9c: Magnitude response in dB of suboptimal methods with r = 5.

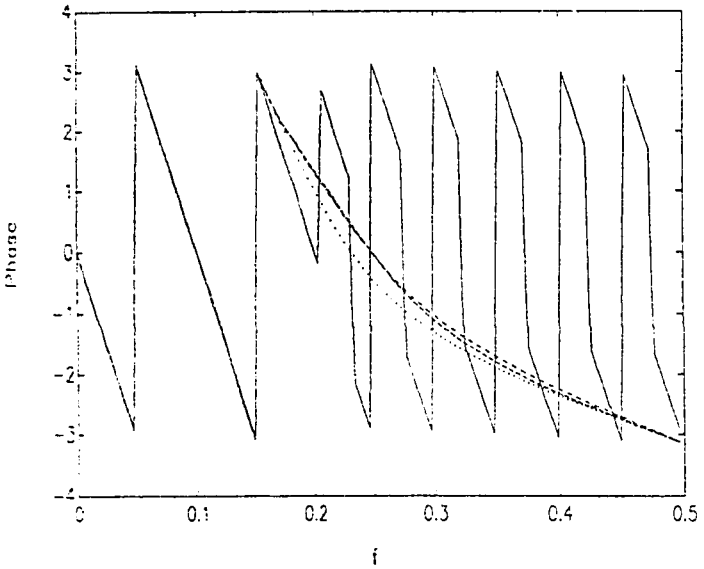


Fig. 5.9d: Phase response of suboptimal methods with r = 5.

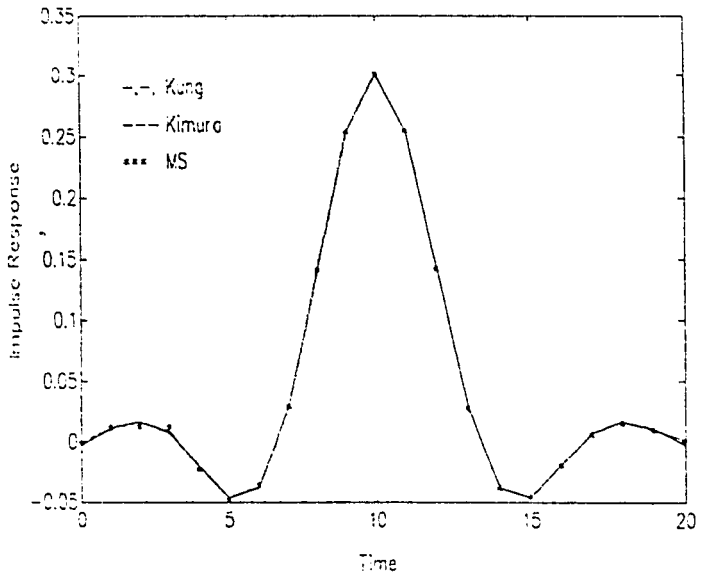


Fig. 5.10a: Impulse response of suboptimal methods with r = 7.

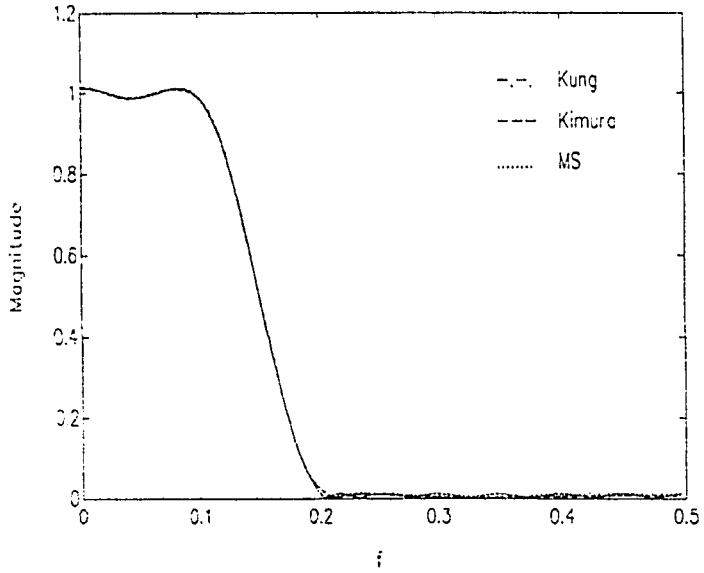


Fig. 5.10b: Magnitude response of suboptimal methods with r = 7.

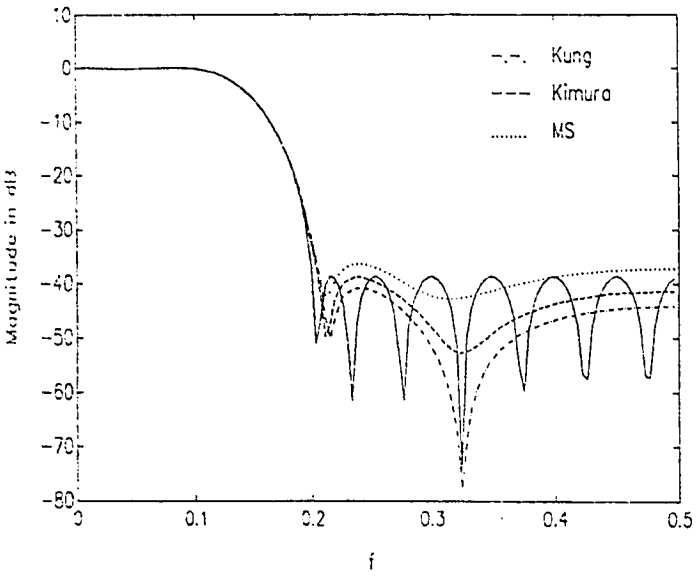


Fig. 5.10c: Magnitude response in dB of suboptimal methods with r = 7.

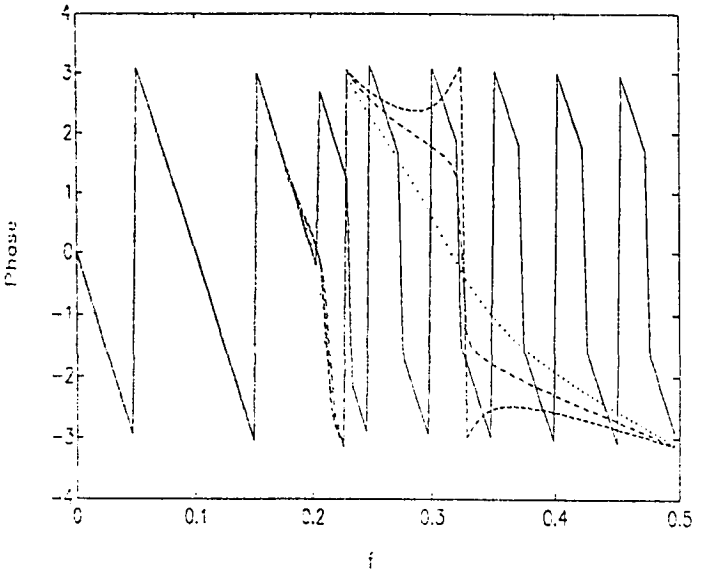


Fig. 5.10d: Phase response of suboptimal methods with r = 7.

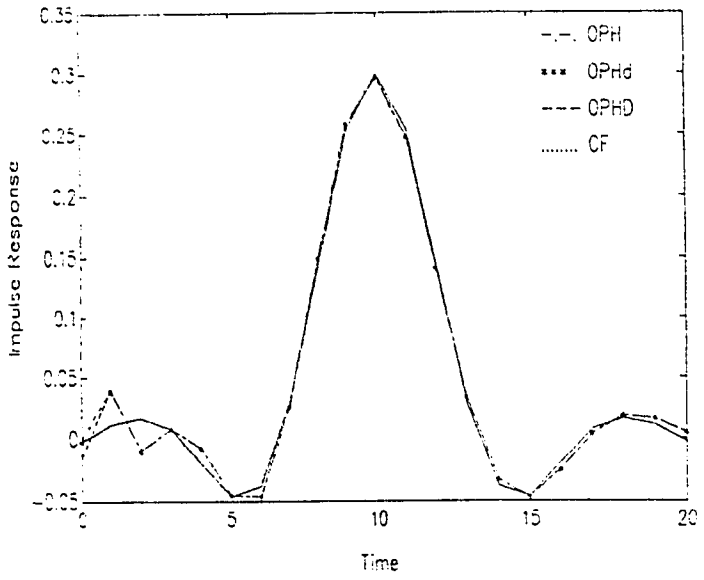


Fig. 5.11a: Impulse response of optimal Hankel methods with r = 5.

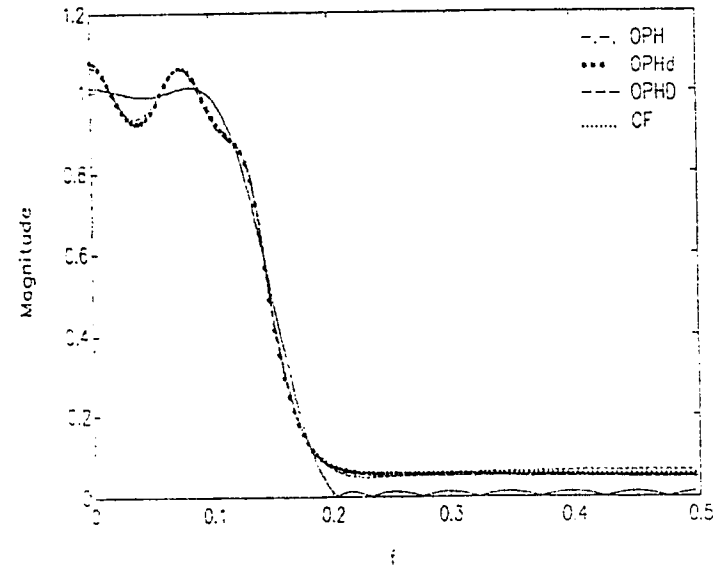


Fig. 5.11b: Magnitude response of optimal Hankel methods with r = 5.

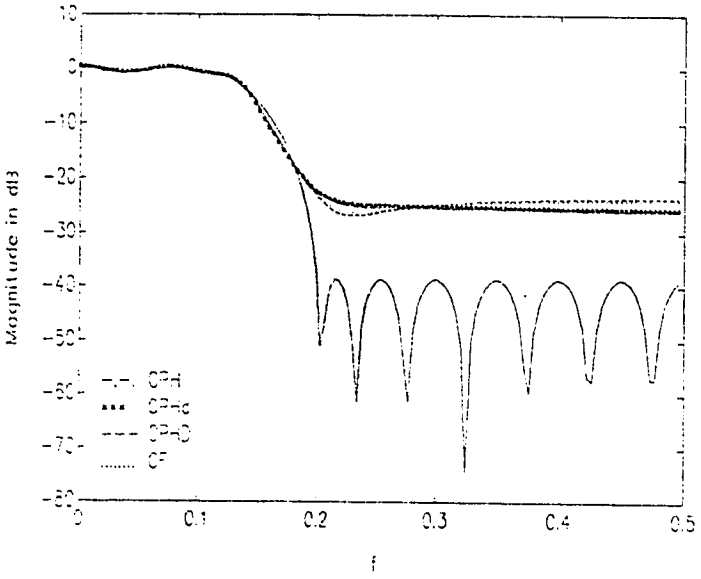


Fig. 5.11c: Magnitude response in dB of optimal Hankel methods with r = 5.

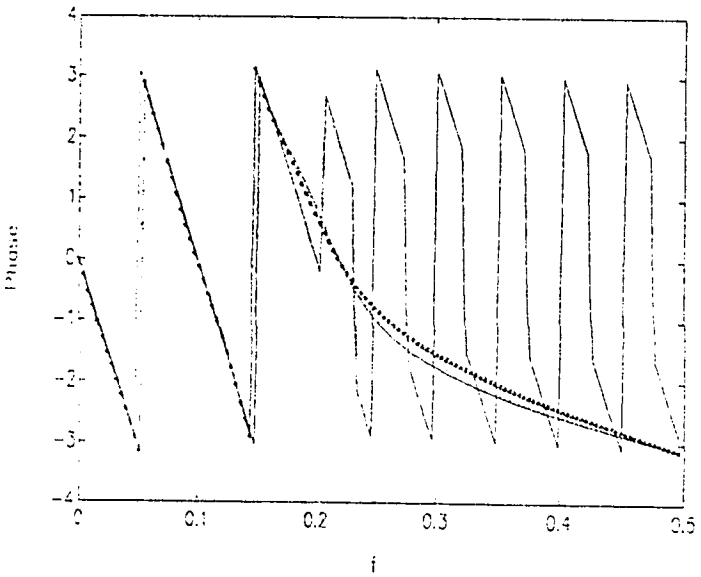


Fig. 5.11d: Phase response of optimal Hankel methods with r = 5.

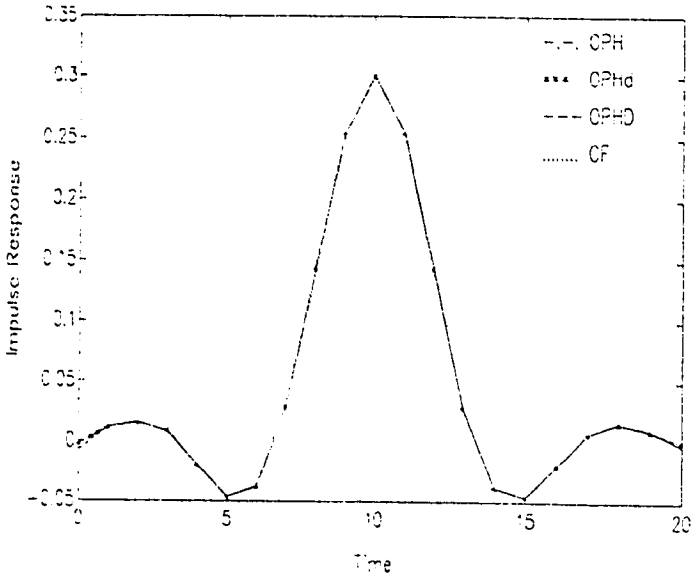


Fig. 5.12a: Impulse response of optimal Hankel methods with r = 7.

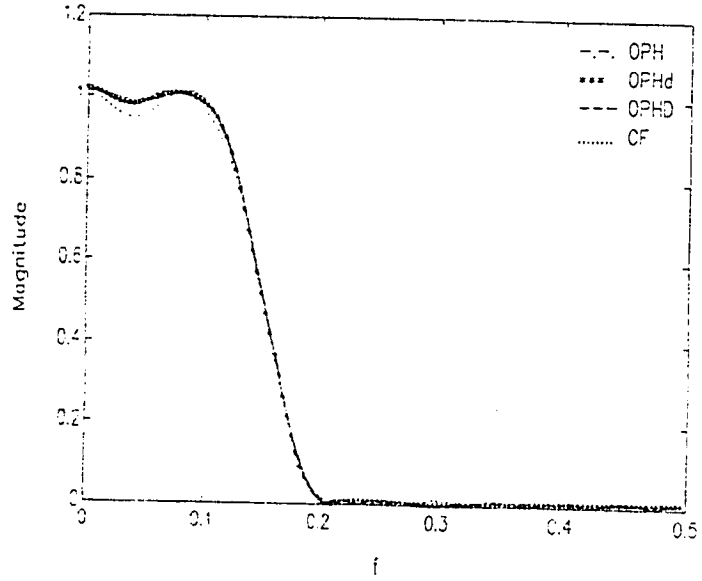


Fig. 5.12b: Magnitude response of optimal Hankel methods with r = 7.

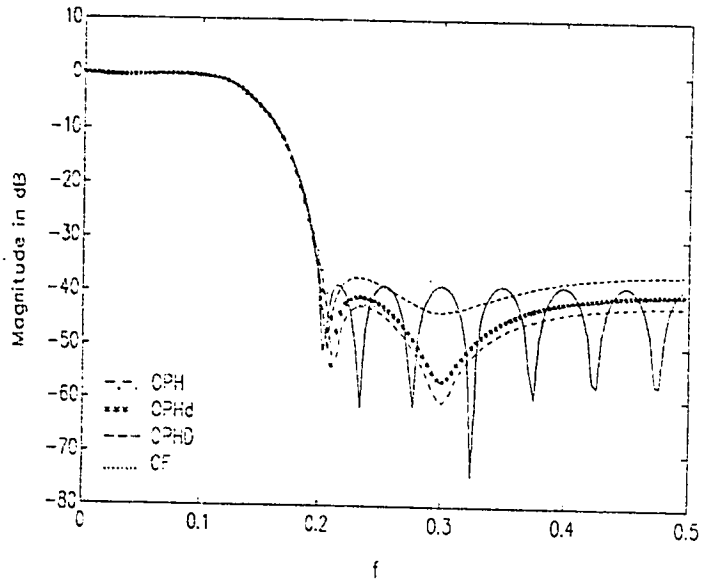


Fig. 5.12c: Magnitude response in dB of optimal Hankel methods with r = 7.

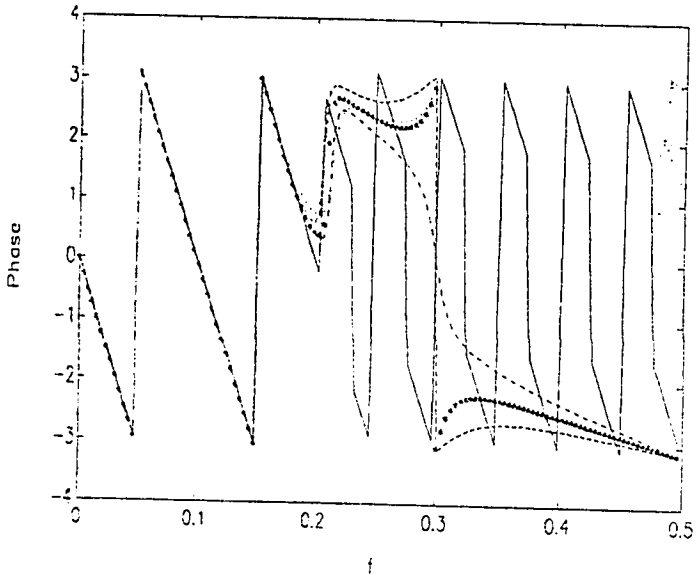


Fig. 5.12d: Phase response of optimal Hankel methods with r =7.

TABLE 7

Error due to Converting the Impulse Response of the CF Method from a Nonparametric Form to a Parametric Form via Prony Method .

	LSE	L_∞
(6,7) IIR filter	0.00221434	0.00540768
(8,8) IIR filter	0.02692528	0.06068403

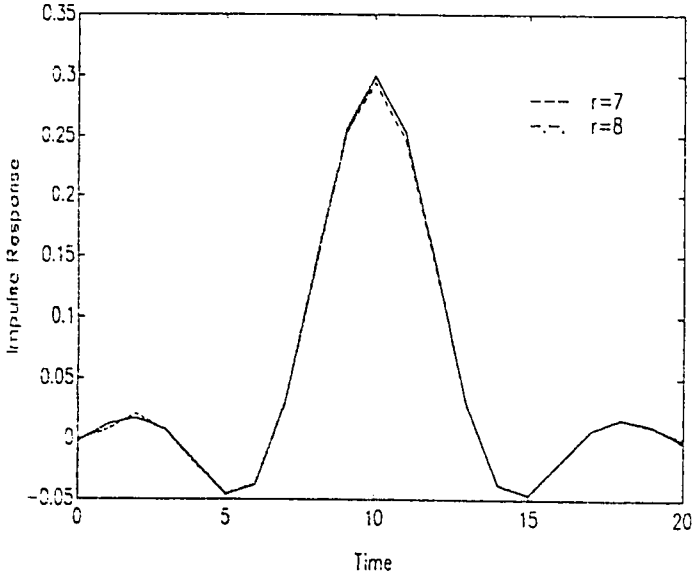


Fig. 5.13a: Impulse responses of (6,7) and (8,8) IIR filters designed using CF method.

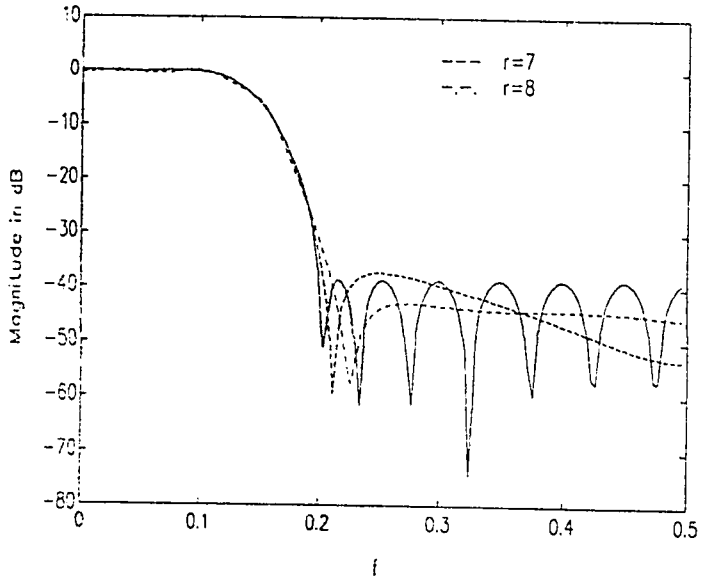


Fig. 5.13b: Magnitude responses in dB of (6,7) and (8,8) IIR filters designed using CF method.

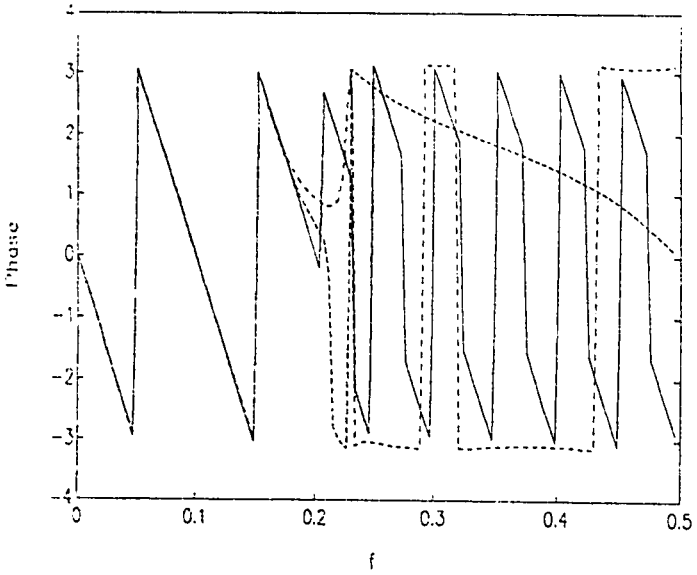


Fig. 5.13c: Phase responses of (6,7) and (8,8) IIR filters designed using CF method.

5.2.3 Example 3: Ideal Differentiator

Consider the ideal differentiator whose characteristic is given by:

$$H(e^{j\omega}) = j\omega e^{-j\omega}, \quad -\pi < \omega < \pi$$

The impulse response is truncated by a Hamming window of length M=57, and shifted to the left by M+1/2 (table 8). Two IIR filters of order 21 and 29 were designed using the above methods to approximate the characteristic of the ideal differentiator.

Only (21,21) IIR filter could be designed using least squares methods. When r = 29, the matrix H^T is almost singular which leads to inaccurate approximation. This is a big limitation of least squares methods. The impulse response, magnitude response, and phase response of (21,21) IIR filter are shown in Fig. 5.14(a,b,c) respectively. Actually, it gives a poor approximation with large error.

When r = 29, Suboptimal methods give an excellent design and the performance of the three methods overlaid with the desired response for both impulse response and frequency response as shown in Fig. 5.16(a,b,c) and table 9. However, the situation is different when r = 21. Kung and Kimura methods give a better design for magnitude response and phase response compared with MS method as shown in Fig. 5.15(b,c). On the other hand, MS method gives a better approximation to the impulse response in general.

The same argument applies for OPH, OPHd, OPHD, and CF methods where optimal IIR filters are obtained when r = 29 (Fig. 5.18(a,b,c)). When

r = 21, OPH , OPHd, and OPHD methods have an almost identical impulse response except at n = 0 which is shown in Fig. 5.17a. Also, they show a similar behaviour in magnitude response. For phase response, the performance of the three methods is comparable at high frequencies but they differ greatly at low frequencies as shown in Fig. 5.17c. Actually, OPH and OPHd methods show identical performance in this example. The CF method gives a poor IIR filter design for r = 21.

It is clear from the above discussion that the performance of suboptimal and optimal Hankel methods is greatly improved when r = 29. This is because of two main factors. The first factor is the difference of order which increased from (21,21) to (29,29). The second factor is the singular value behaviour of the Hankel matrix given in table 8. While $\sigma_{29} >> \sigma_{30}$, the separation between σ_{21} and σ_{22} is small.

TABLE 8

Impulse Response and Singular Values of Example 3

Impulse response		Singular values	
- 0.00285714	- 0.99710760	2.83883692	0.05456683
0.00307009	0.49423342	2.80665962	0.00122220
- 0.00352051	- 0.32472878	2.66301335	0.00011836
0.00423252	0.23861142	2.59495146	0.00003435
- 0.00523143	- 0.18589863	2.45416349	0.00002380
0.00654378	0.14994041	2.41935101	0.00001119
- 0.00819807	- 0.12360987	2.29932511	0.00000688
0.01022528	0.10335066	2.23841053	0.00000569
- 0.01265973	- 0.08719275	2.11599299	0.00000523
0.01554028	0.07395865	2.02103310	0.00000493
- 0.01891186	- 0.06290258	1.90506116	0.00000398
0.02282774	0.05352997	1.80317893	0.00000398
- 0.02735252	- 0.04550028	1.68928543	0.00000388
0.03256642	0.03857143	1.58525294	0.00000386
- 0.03857143	- 0.03256642	1.47210972	0.00000335
0.04550028	0.02735252	1.36761310	0.00000309
- 0.05352997	- 0.02282774	1.25438155	0.00000301
0.06290258	0.01891186	1.15171489	0.00000297
- 0.07395865	- 0.01554028	1.03639609	0.00000272
0.08719275	0.01265973	0.99044438	0.00000266
- 0.10335066	- 0.01022528	0.91720400	0.00000256
0.12360987	0.00819807	0.81828003	0.00000253
- 0.14994041	- 0.00654378	0.70771332	0.00000243
0.18589863	0.00523143	0.60009676	0.00000242
- 0.23861142	- 0.00423255	0.49054970	0.00000236
0.32472878	0.00352051	0.38188384	0.00000236
- 0.49423342	- 0.00307009	0.27264990	0.00000113
0.99710762	0.00285714	0.16367285	0.00000002
0			

TABLE 9

LSE and L_∞ Error Norms of all Methods Applied to Example 3.

		r = 21		r = 29	
		LSE	L_∞	LSE	L_∞
Least squares methods	Shank	1.71132729	1.00975380		
	Prony	1.72775204	2.97898270		
	Pade	$8.47015857 \times 10^{21}$	2.94540074		
Suboptimal methods	Kung	0.43087535	1.33548011	0.00285807	0.00475097
	Kimura	0.43086587	1.33292295	0.00007264	0.00193592
	MS	0.50121236	1.16201010	0.00012964	0.00224881
Optimal Hankel methods	OPH	0.49787267	1.04121108	0.00285909	0.00407906
	OPHd	0.49786448	1.04131641	0.00010549	0.00122245
	OPHD	0.59345225	1.20165014	0.00010550	0.00122233
	CF	1.70275005	2.84526949	0.00010549	0.00122243

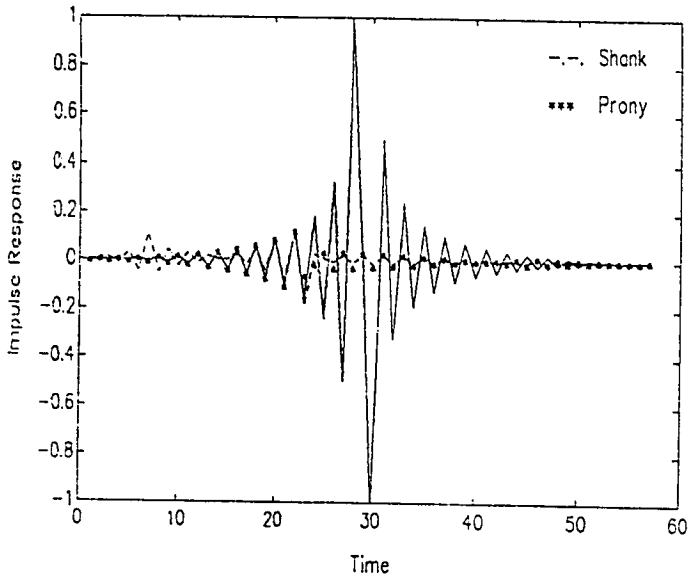


Fig. 5.14a: Impulse response of least squares methods with r = 21.

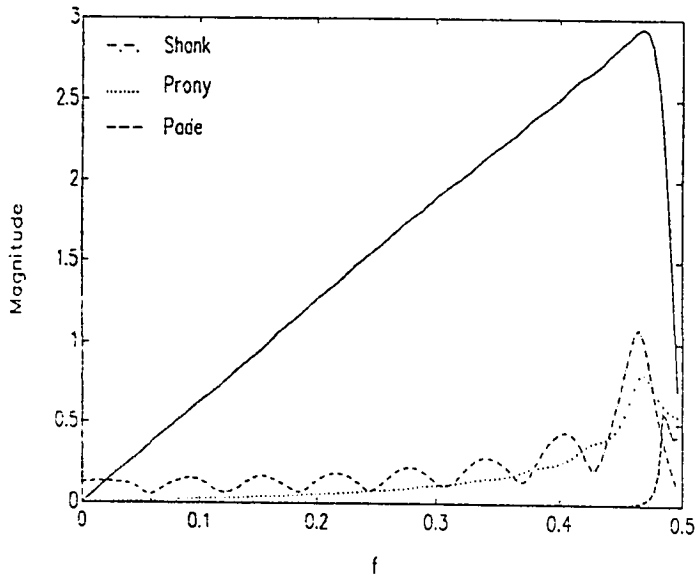


Fig. 5.14b: Magnitude response of least squares methods with r = 21.

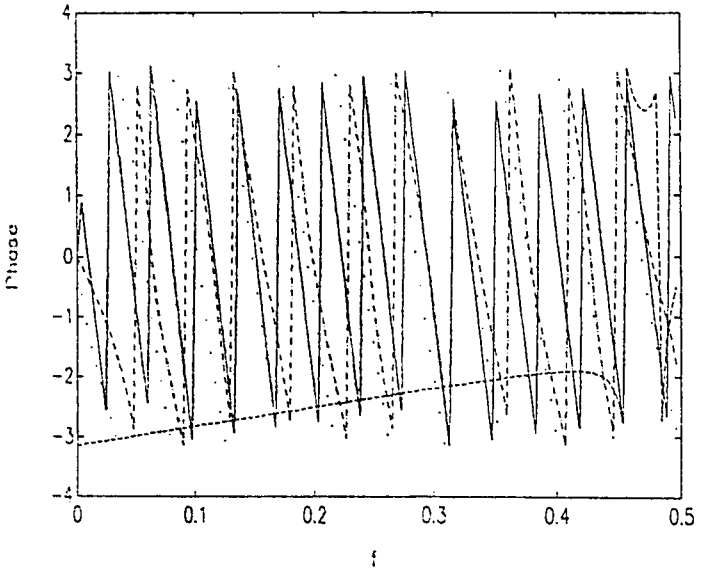


Fig. 5.14c: Phase response of least squares methods with r = 21.

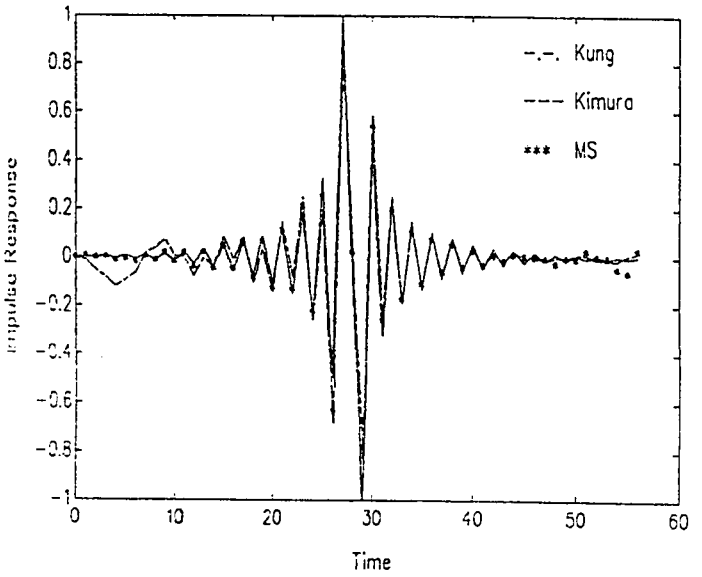


Fig. 5.15a: Impulse response of suboptimal methods r = 21.

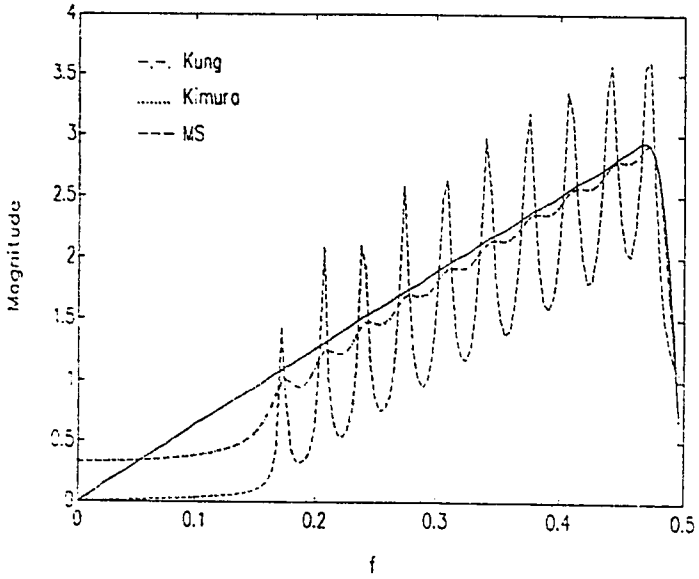


Fig. 5.15b: Magnitude response of suboptimal methods with r = 21.

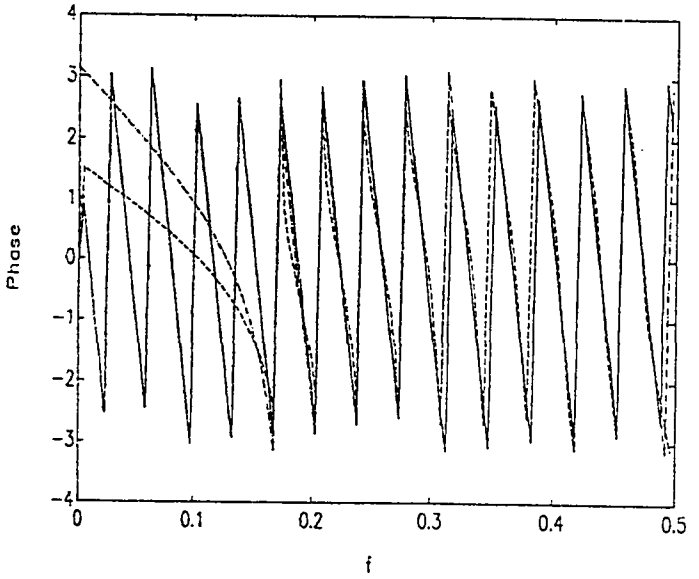


Fig. 5.15c: Phase response of suboptimal methods with r = 21.

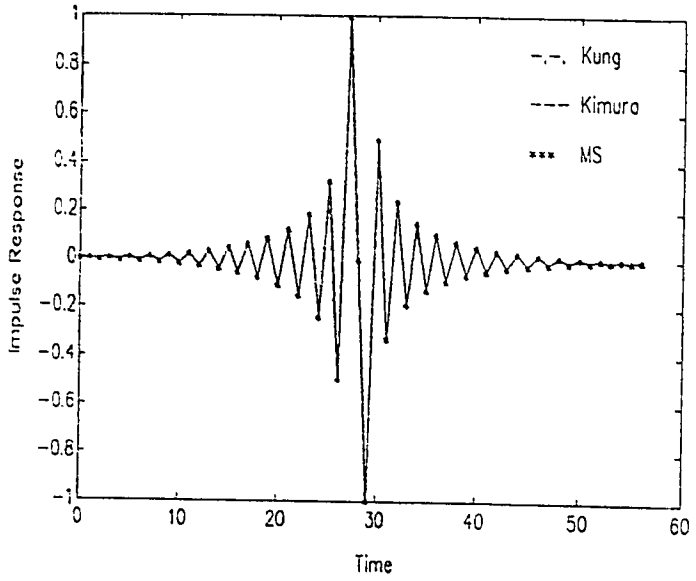


Fig. 5.16a: Impulse response of suboptimal methods with r = 29.

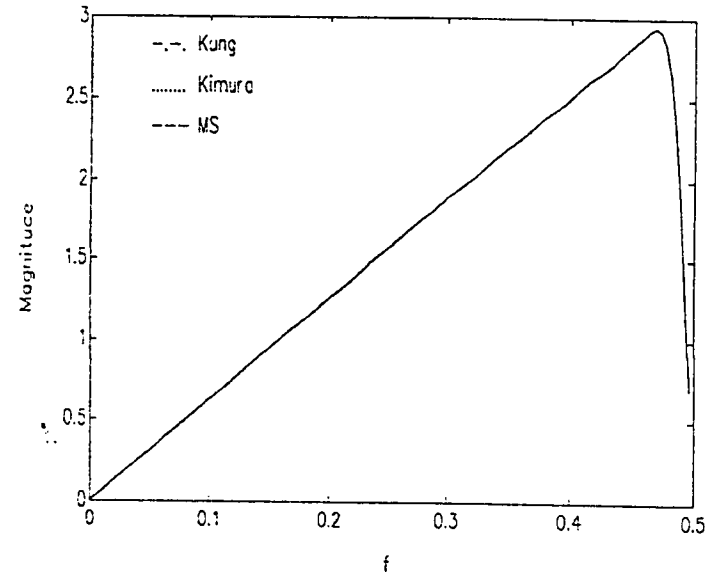


Fig. 5.16b: Magnitude response of suboptimal methods with r = 29.

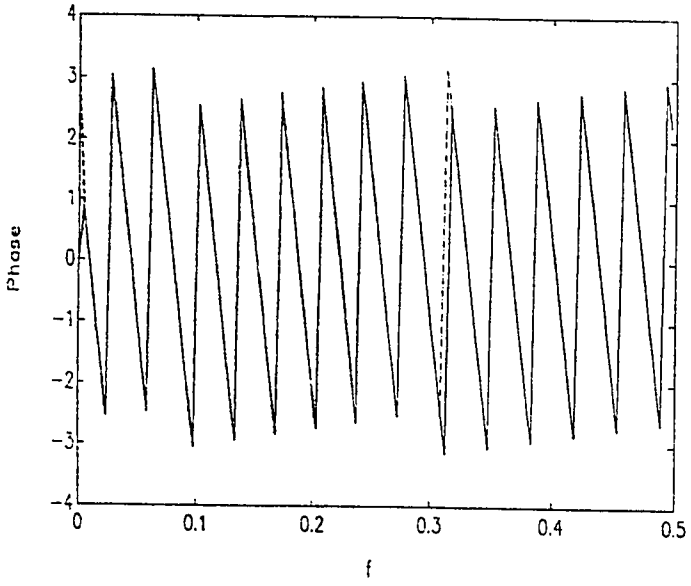


Fig. 5.16c: Phase response of suboptimal methods with r = 29.

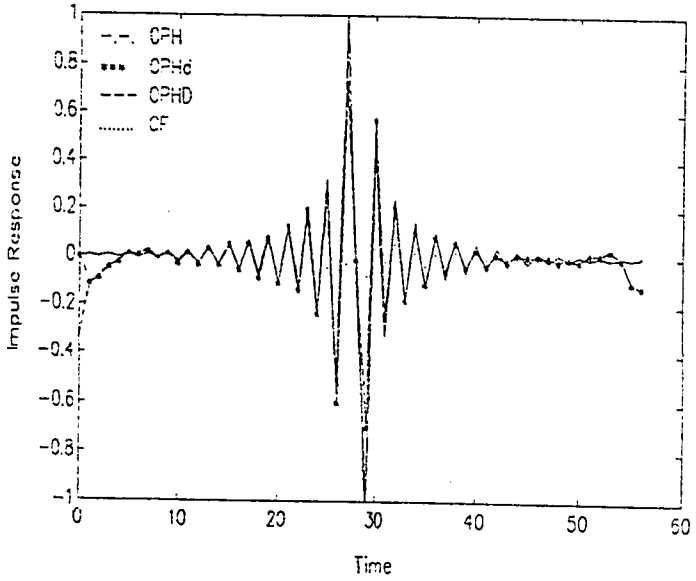


Fig. 5.17a: Impulse response of optimal Hankel methods with r = 21.

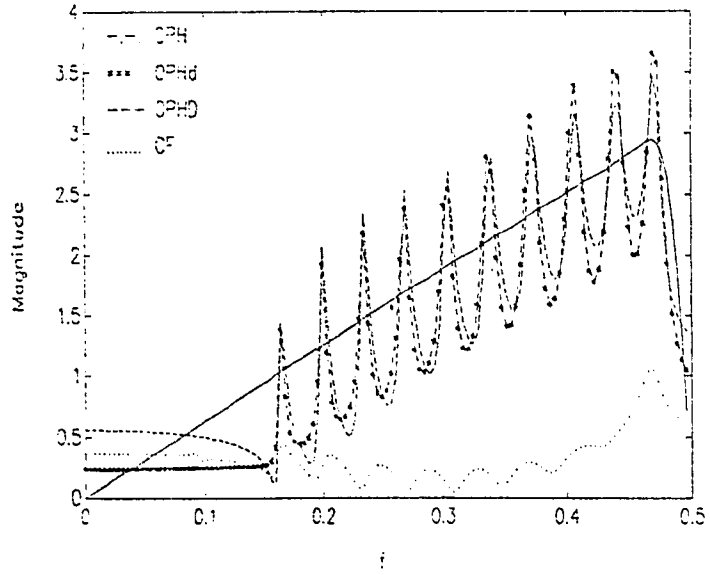


Fig. 5.17b: Magnitude response of optimal Hankel methods with r = 21.

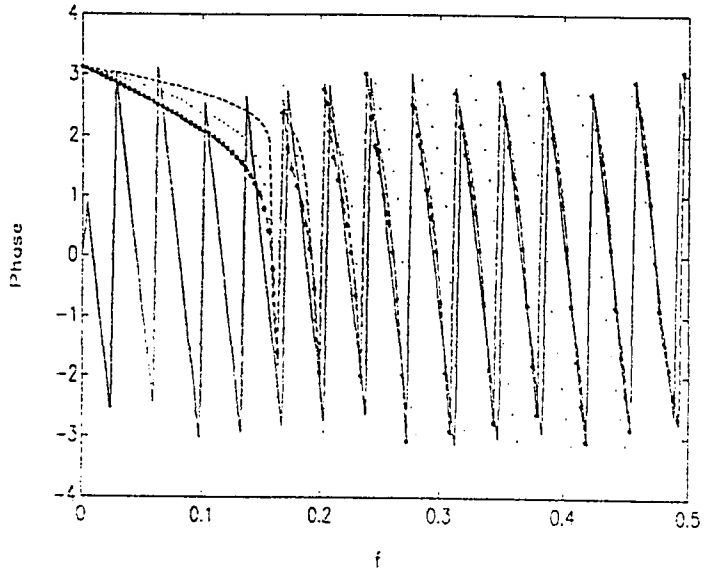


Fig. 5.17c: Phase response of optimal Hankel methods with r = 21.

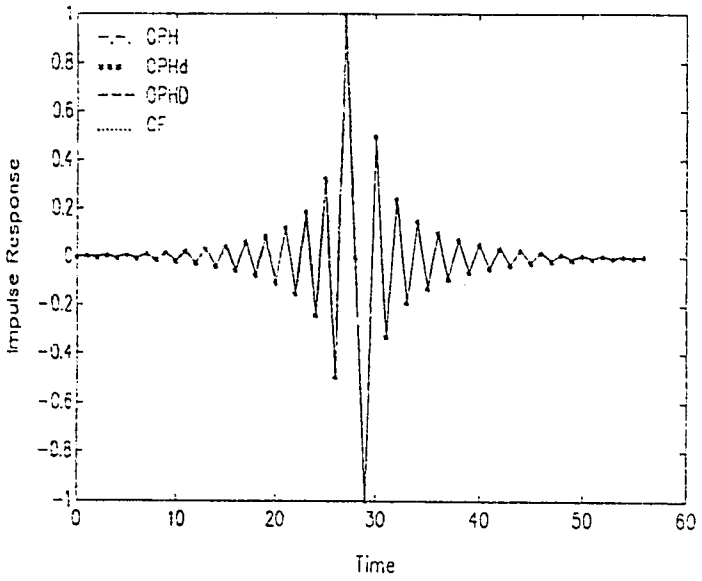


Fig. 5.18a: Impulse response of optimal Hankel methods with r = 29.

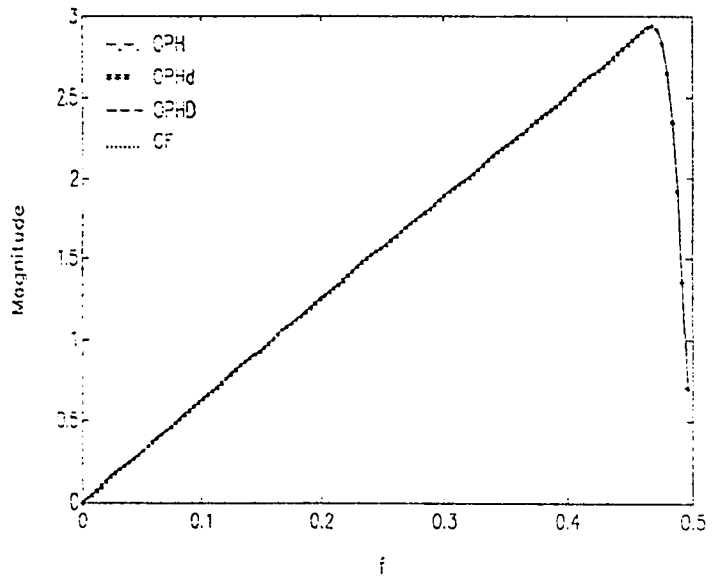


Fig. 5.18b: Magnitude response of optimal Hankel methods with r = 29.

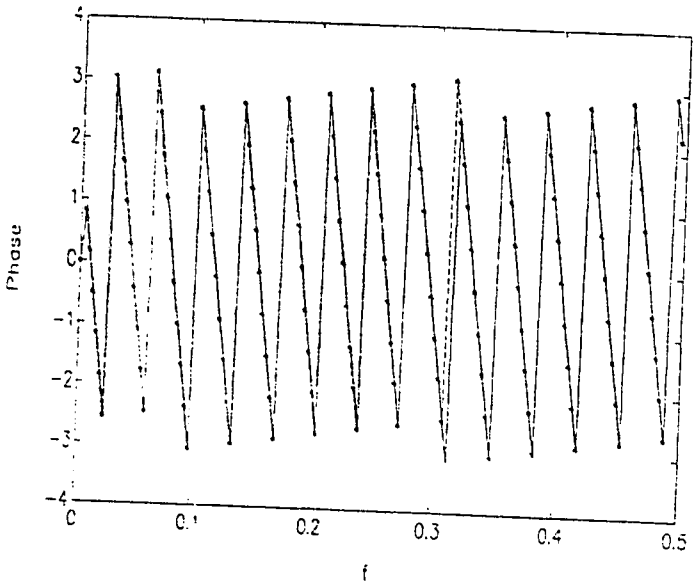


Fig. 5.18c: Phase response of optimal Hankel methods with r = 29.

5.2.4 Example 4: Ideal BPF

An ideal HPF is to be designed with $f_c = 0.35$. The impulse response of the ideal filter is windowed using Hamming window of length n = 45. The windowed impulse response is shown in table 10. Two IIR filters of order 9 and order 11 were designed for simulation.

The results obtained using least squares methods indicate the weakness of these techniques specially Pade approximation. The impulse response of the (11,11) IIR filter (the same applies for r = 9) designed using Pade approximation is shown in Fig. 5.19. It becomes totally unstable after n > 30. The first M+N+1 samples of the impulse response is shown in table 12 which shows the efficiency of Pade technique in approximating the impulse response over this range. The impulse response obtained using shank and Prony methods also suffer from a serious error when r = 9 as shown in Fig. 5.20a. For r = 11, Shank method gives a good approximation to the impulse response in general which is not the case for Prony method. The three methods show a very bad performance in magnitude response. The phase response of Shank and Prony methods show a poor performance at low frequencies but at high frequencies the performance is acceptable specially for Shank method. This is shown in Fig. 5.20(b,c) and Fig. 5.21(b,c).

The performance of the IIR filters designed using suboptimal methods is shown in Fig. 5.22(a,b,c) for r = 9 and in Fig. 5.23(a,b,c) for r = 11. They clearly indicate the optimality of these methods compared to least squares methods. The impulse response of the three methods almost overlaid with each other for both orders. For magnitude response, they give an almost exact

approximation when r = 11. However, Kung and Kimura methods are slightly better at high frequencies. This difference is apparent when r = 9 where MS method gives a better magnitude response at low frequencies and Kung and Kimura methods have a better approximation at high frequencies. For phase response, the three methods show a linear phase behaviour at high frequencies. The situation is different at low frequencies where the phase response is nearly linear for Kung and Kimura methods and nonlinear for MS method.

Optimal Hankel methods and CF method give an excellent IIR filter design when r = 11. The impulse responses and magnitude responses of the four methods almost overlaid with the desired response. For the phase response, the four methods give an exact matching with the desired response at the pass-band but this is not the case at the stop-band. This is shown in Fig. 5.25(a,b,c). The most important observation is the bad performance of the CF method when r = 9 as shown in Fig . 5.24(a,b,c). Finally, OPH and OPHd methods have a better performance at low frequencies for both orders.

TABLE 10

Impulse Response and Singular Values of Example 4.

Impulse response		Singular values	
0.00109905	- 0.25589667	0.99837873	0.00056531
- 0.00103675	0.14830400	0.99837170	0.00054655
0.00000000	- 0.03137379	0.99813938	0.00052923
0.00164503	- 0.04328851	0.99437534	0.00051475
- 0.00256943	0.05641757	0.96533676	0.00050156
0.00111117	- 0.02618969	0.85453759	0.00049054
0.00278749	- 0.01106467	0.62604683	0.00048070
- 0.00617166	0.02762058	0.35697537	0.00047255
0.00465530	- 0.01912819	0.16228746	0.00046549
0.00310024	- 0.00000000	0.06262465	0.00045979
- 0.01195195	0.01262129	0.02220782	0.00045513
0.01262129	- 0.01195195	0.00776540	0.00045162
- 0.00000000	0.00310024	0.00280920	0.00044911
- 0.01912819	0.00465530	0.00119148	0.00044763
0.02762059	- 0.00617166	0.00082348	0.00007294
- 0.01106467	0.00278749	0.00077570	0.00005936
- 0.02618969	0.00111117	0.00074151	0.00000341
0.05641757	- 0.00256943	0.00070578	0.00000216
- 0.04328851	0.00164503	0.00067080	0.00000066
- 0.03137379	0.00000000	0.00064054	0.00000040
0.14830400	- 0.00103675	0.00061182	0.00000016
- 0.25589667	0.00109905	0.00058777	0.00000002
0.29951344			

TABLE 11

LSE and L_∞ Error Norms of all Methods Applied in Example 4.

		r = 9		r = 11	
		LSE	L_∞	LSE	L_∞
Least squares methods	Shank	0.33378227	1.91227702	0.05281619	2.68778652
	Prony	0.54533590	1.54112088	0.50623604	2.11186610
	Pade	4.19210282×10^{8}	1.00038454	1.01274028×10^{3}	1.00304463
Suboptimal methods	Kung	0.03412645	0.11674964	0.00434937	0.01412746
	Kimura	0.03410875	0.11596599	0.00420822	0.01416025
	MS	0.03066517	0.12479650	0.00369680	0.01582372
Optimal Hankel methods	OPH	0.05055278	0.07511337	0.00629298	0.00896005
	OPHd	0.05054084	0.07620707	0.00619627	0.00870955
	OPHD	0.05175175	0.06756902	0.00619669	0.00863803
	CF	0.23325208	0.69991663	0.00624079	0.00957150

TABLE 12

The First M+N+1 Samples of the Impulse Response of (11,11) IIR Filter Designed Using Pade Approximation.

Desired	Pade
0.00109905347452	0.00109905347452
- 0.00103675311638	- 0.00103675311638
0.00000000000000	0.00000000000000
0.00164502721676	0.00164502721676
- 0.00256942998681	- 0.00256942998681
0.00111117524183	0.00111117524183
0.00278748702126	0.00278748702126
- 0.00617166302898	- 0.00617166302897
0.00465530490147	0.00465530490146
0.00310023600453	0.00310023600456
- 0.01195194761291	- 0.01195194761297
0.01262129480804	0.01262129480817
- 0.00000000000000	- 0.00000000000023
- 0.01912818708308	- 0.01912818708271
0.02762058751401	0.02762058751342
- 0.01106467036706	- 0.01106467036614
0.02618969521210	0.02618969521210
0.05641757342656	0.05641757342901
- 0.04328851311492	- 0.04328851311892
- 0.03137379291959	- 0.03137379291329
0.14830400144958	0.14830400144013
- 0.25589667033296	- 0.25589667031949
0.29951344282816	0.29951344280967

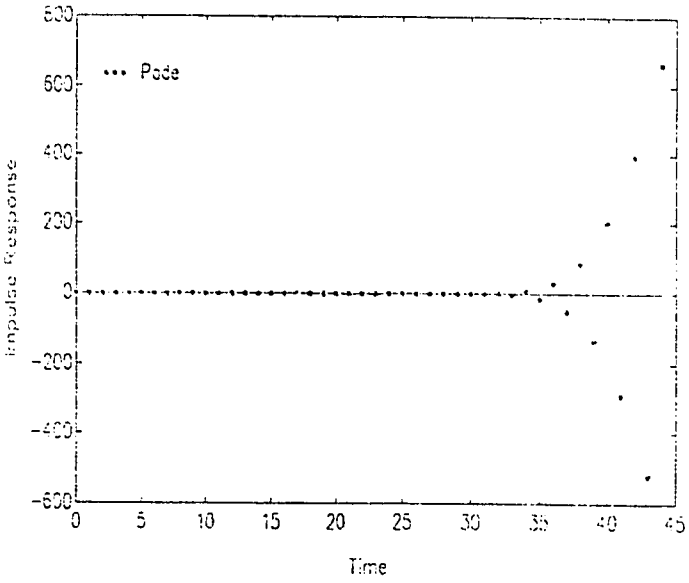


Fig. 19: Impulse response of Pade approximation method with r = 11.

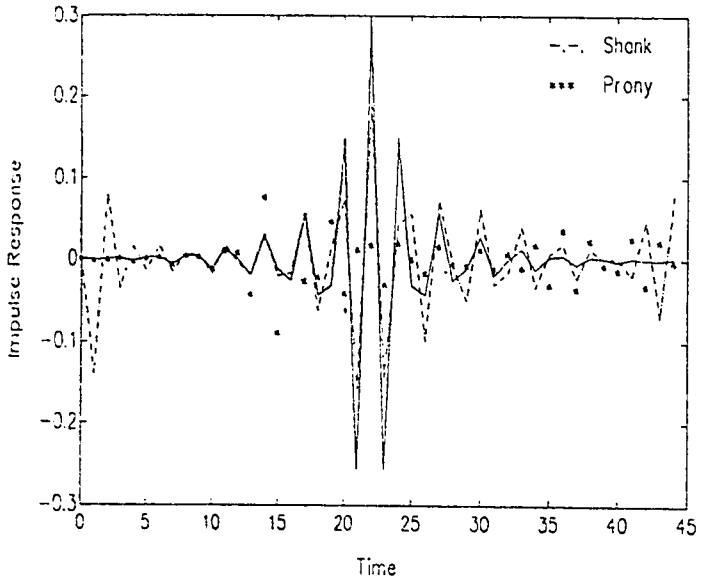


Fig. 5.20a: Impulse response of least squares methods with r = 9.

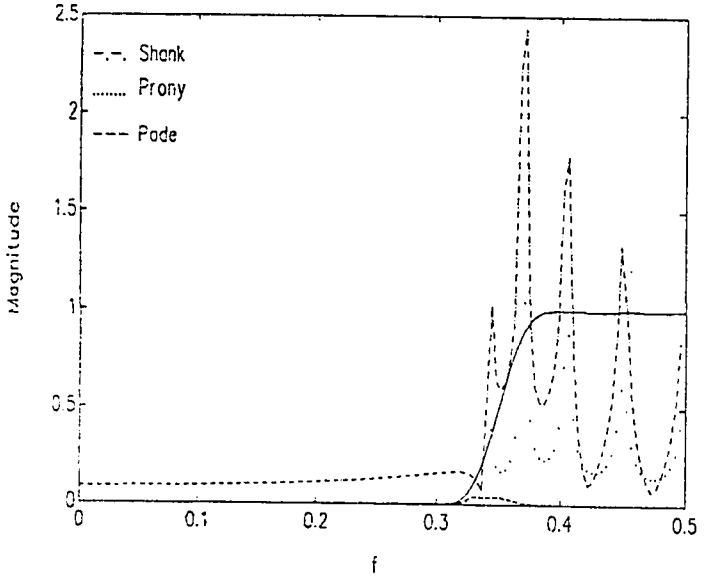


Fig. 5.20b: Magnitude response of least squares methods with r = 9.

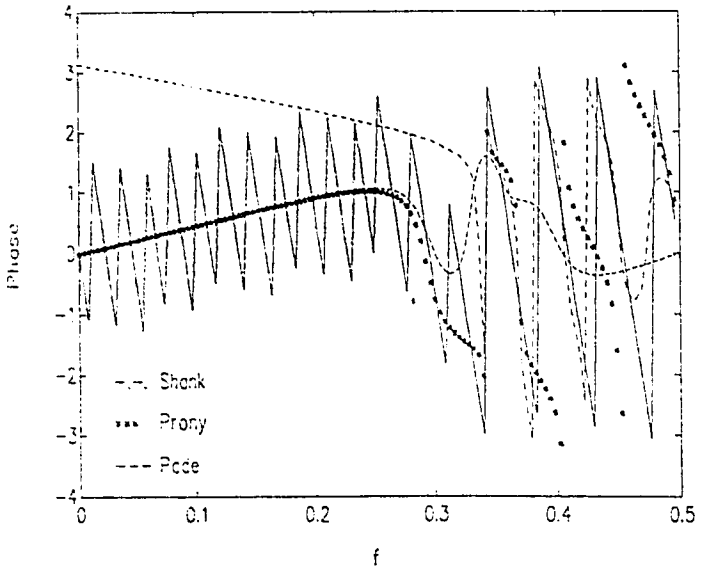


Fig. 5.20c: Phase response of least squares methods with r = 9.

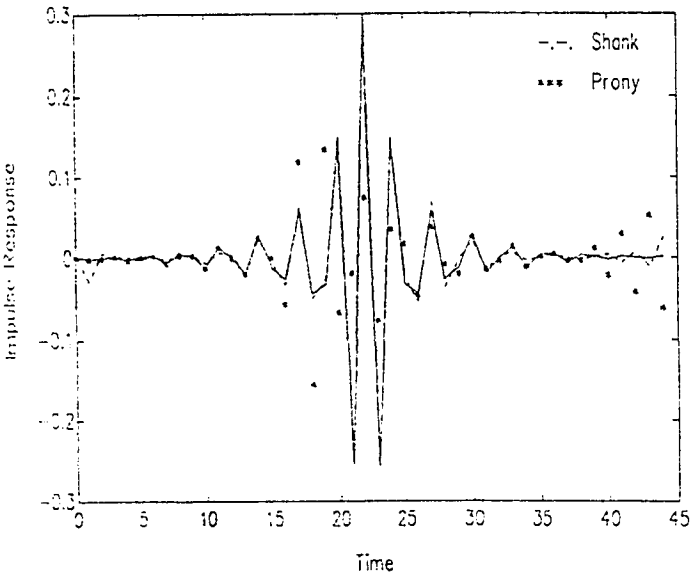


Fig. 5.21a: Impulse response of least squares methods with r = 11.

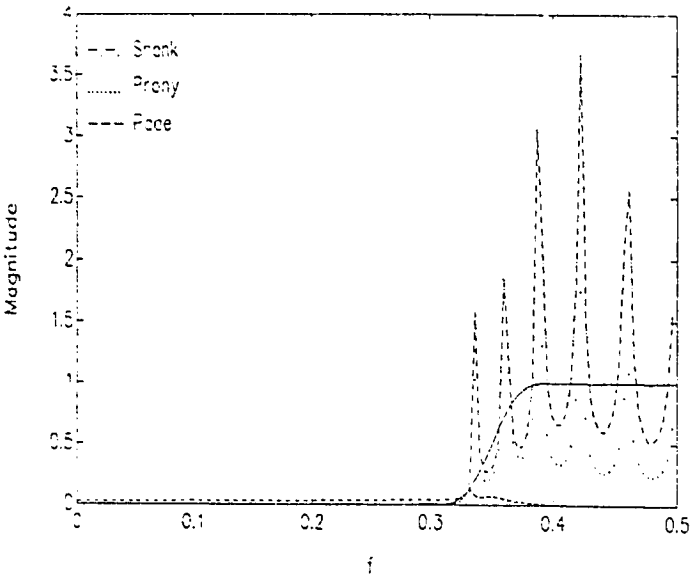


Fig. 5.21b: Magnitude response of least squares methods with r = 11.

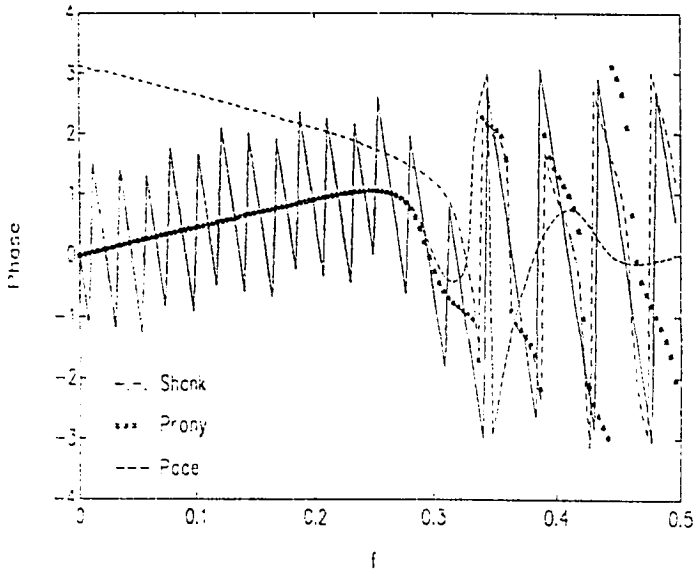


Fig. 5.21c: Phase response of least squares methods with r = 11.

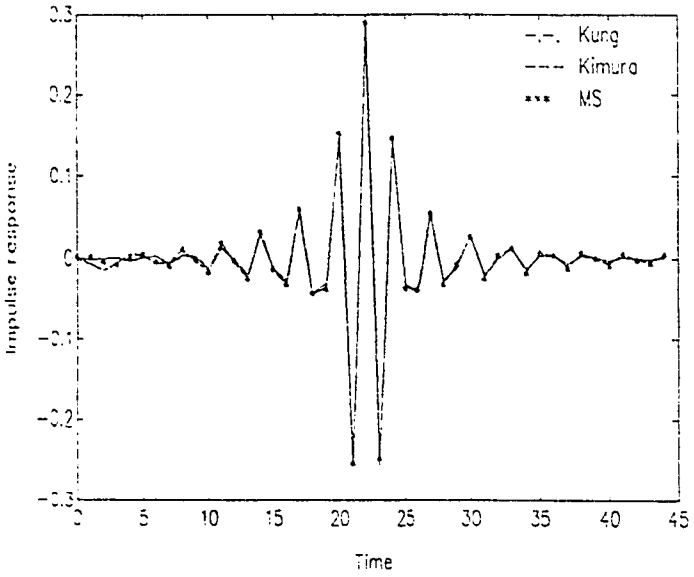


Fig. 5.22a: Impulse response of suboptimal methods with r = 9.

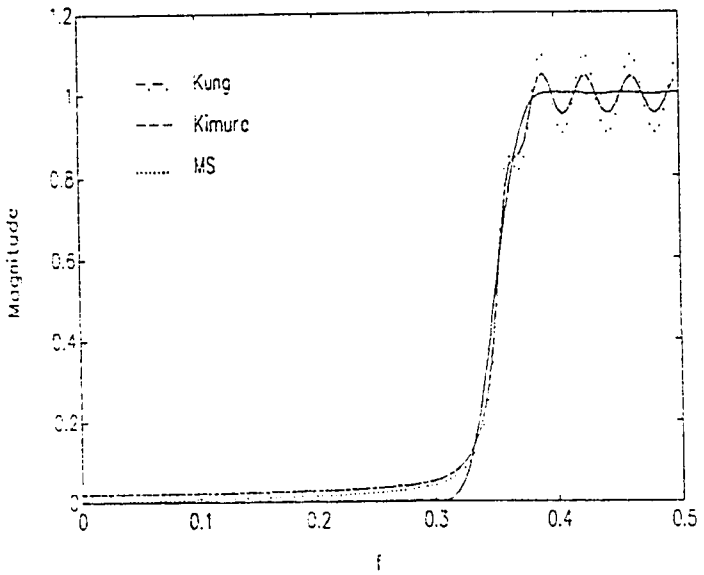


Fig. 5.22b: Magnitude response of suboptimal methods with r = 9.

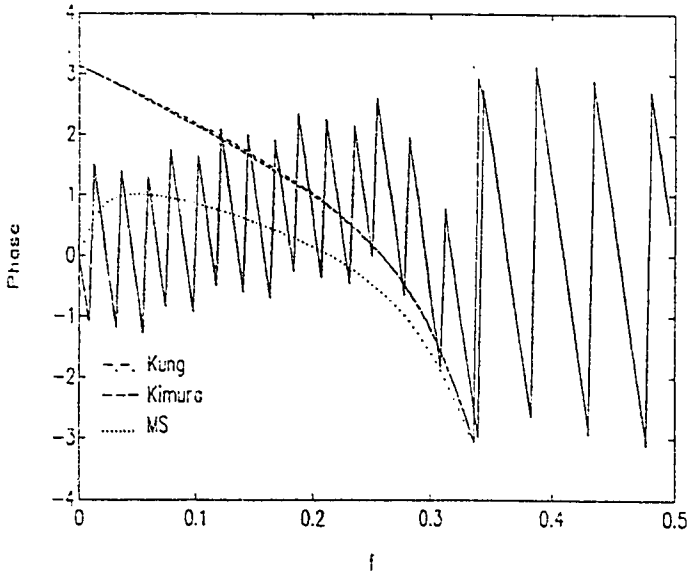


Fig. 5.22c: Phase response of suboptimal methods with r = 9.

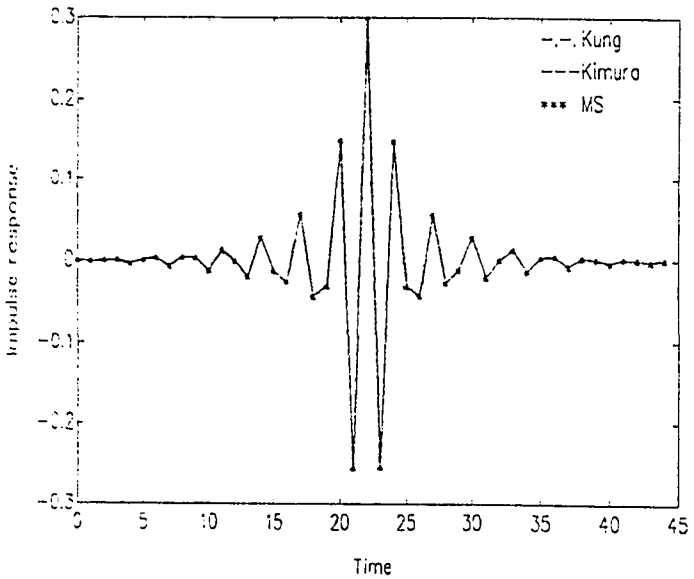


Fig. 5.23a: Impulse response of suboptimal methods with r = 11.

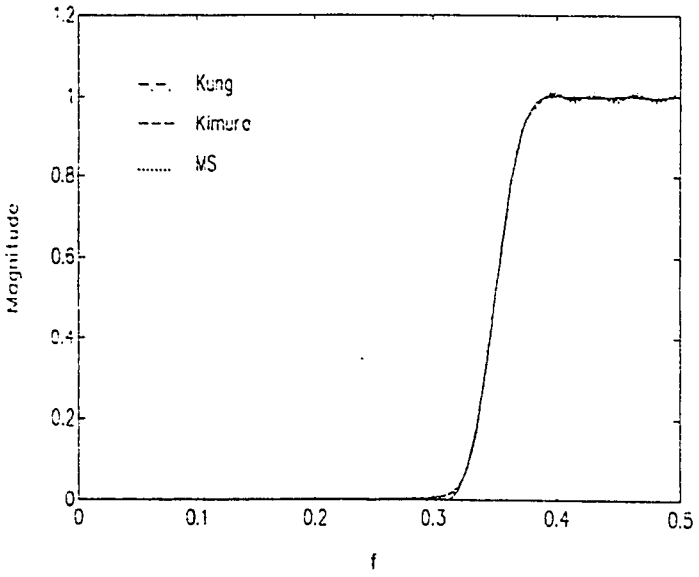


Fig. 5.23b: Magnitude response of suboptimal methods with r = 11.

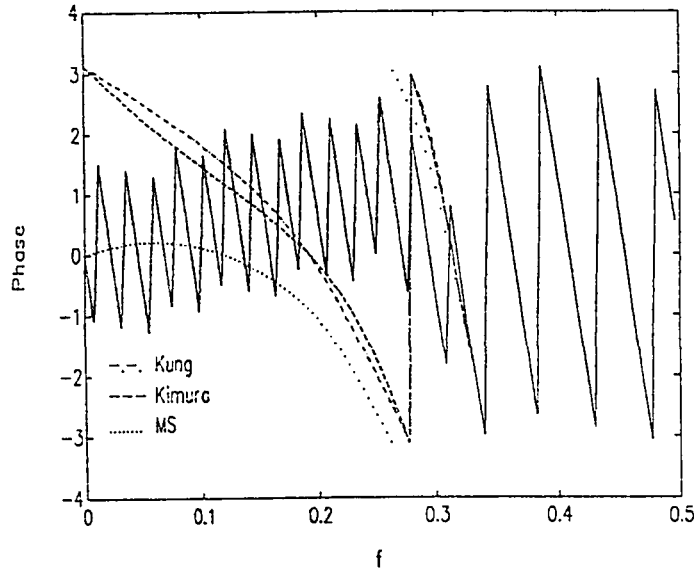


Fig. 5.23c: Phase response of suboptimal methods with r = 11.

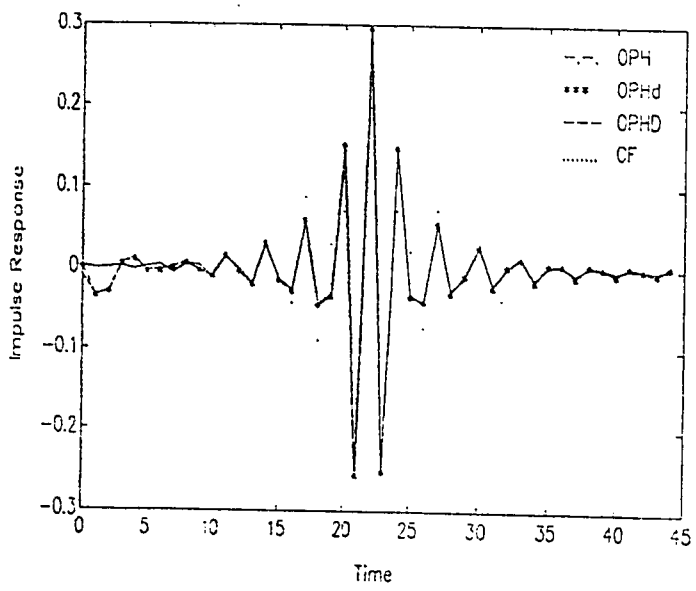


Fig. 5.24a: Impulse response of optimal Hankel methods with r = 9.

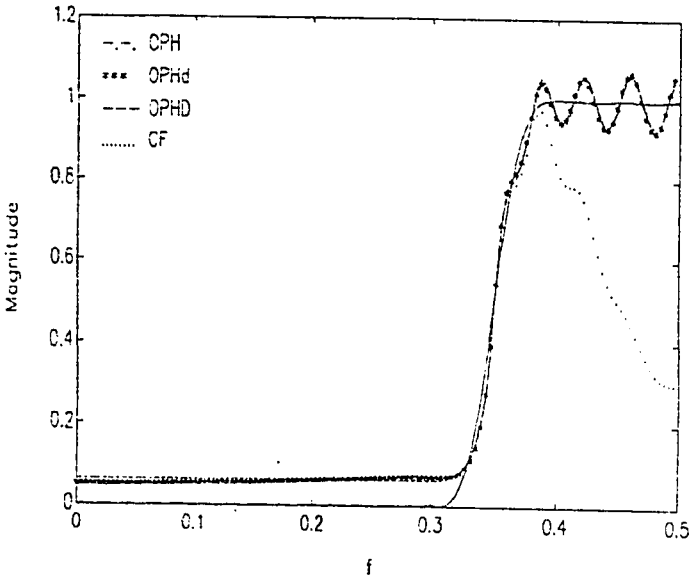


Fig. 5.24b: Magnitude response of optimal Hankel methods with r = 9.

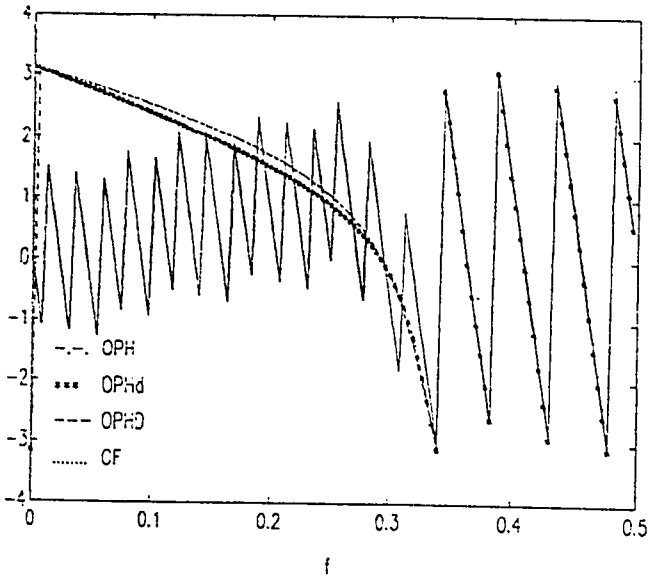


Fig. 5.24c: Phase response of optimal Hankel methods with r = 9.

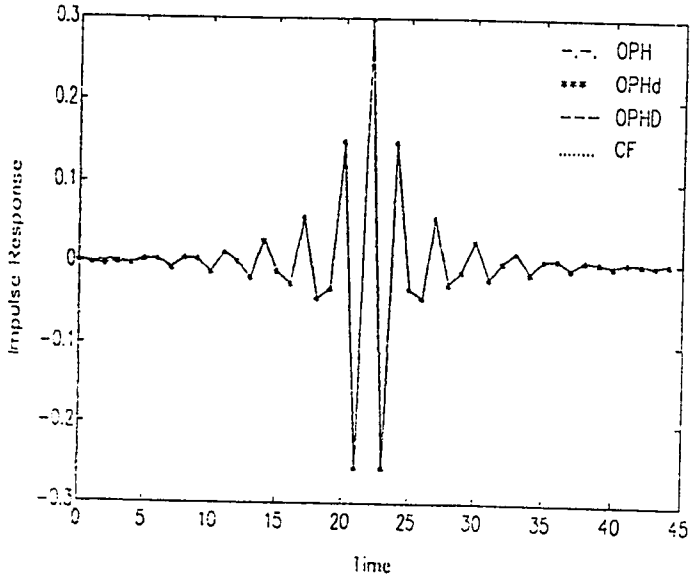


Fig. 5.25a: Impulse response of optimal Hankel methods with r = 11.

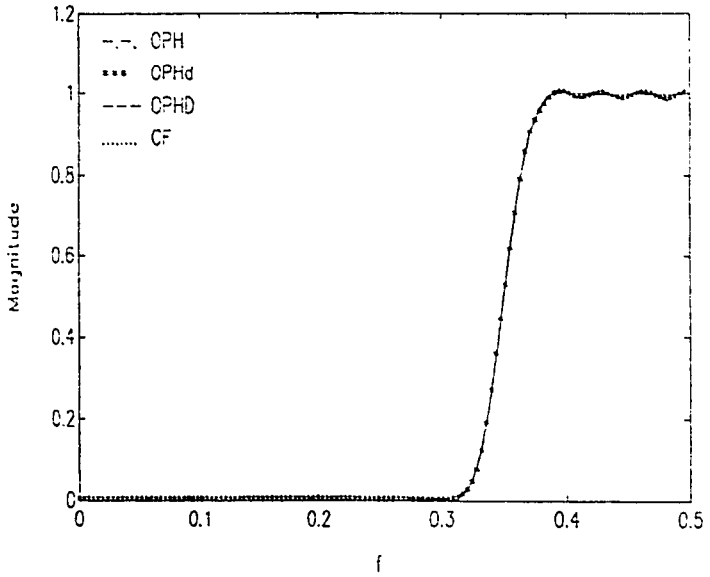


Fig. 5.25b: Magnitude response of optimal Hankel methods with r = 11.

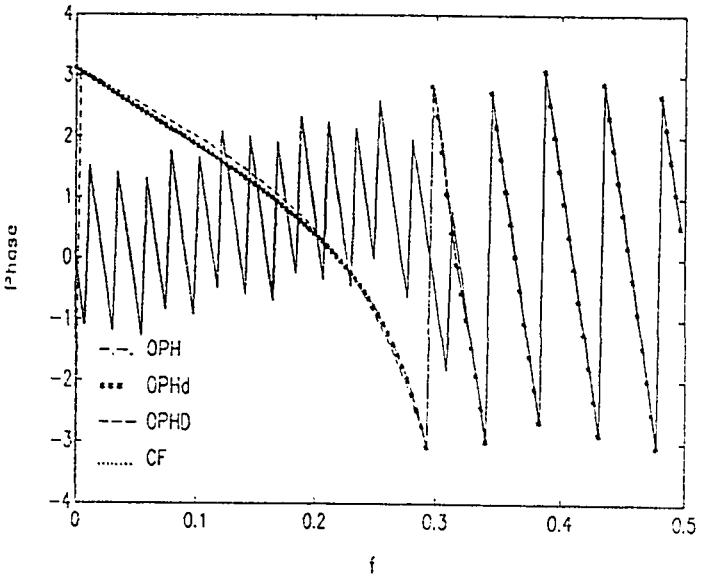


Fig. 5.25c: Phase response of optimal Hankel methods with r = 11.

5.2.5 Example 5: Ideal HPF

An ideal BPF filter with $f_{c1} = 0.15$ and $f_{c2} = 0.30$ is to be approximated by a (12,12) and (16,16) IIR filters. The windowed impulse response of length 51 is shown in table 13.

Least squares methods show a very poor performance in frequency response as shown in Fig. 5.26(b,c) for r = 12 and Fig. 5.27(b,c) for r = 16. However, when r =16, they give a very efficient approximation to the impulse response except for Pade approximation method. This is shown in Fig. 5.27a. When r =12, Shank method and Prony method have a rather large LSE compared to the other methods which could be seen from table 14. Note that the impulse response obtained by Shank method is comparatively better than the one obtained by Prony method in general (Fig. 5.26a)

Suboptimal methods performance is shown in Fig. 5.28(a,b,c) and Fig. 5.29(a,b,c). For r = 12, the performance of the three methods is comparable. However, while Kung and Kimura methods show a better approximation to the magnitude response of the desired filter at mid-band frequencies, MS method has a better attenuation at low frequencies. For phase response, these methods show a linear phase characteristics in general but MS method deviates from linearity at low frequencies. When r = 16, suboptimal methods give an almost a perfect matching to the desired response except for phase response which deviate slightly from linearity at low and high frequencies.

Optimal Hankel methods show also an excellent approximation to the ideal BPF when r = 16 as shown in Fig. 5.31(a,b,c). The impulse response and magnitude response of the four methods almost overlaid with the desired

response. For phase response, these methods give a perfect matching at mid-band frequencies but they deviate slightly at low and high frequencies. Note that the phase response of the OPH method has a better linearity at low and high frequencies. When r = 12, the performance of the three methods is highly comparable. For magnitude response, these methods show a similar behaviour at mid-band frequencies. However, OPH, OPHd, and CF methods have a better attenuation at low frequencies compared to OPHD. In contrast, OPHD has more attenuation at high frequencies as shown Fig. 5.30b. The phase response is totally linear at mid-band frequencies but they somewhat diverge from linearity at low and high frequencies which is shown in Fig. 5.30c. Note that the performance of OPH method and OPHd method is almost identical for this example since the D-term is very small in this example.

Finally, suboptimal methods (suboptimal Hankel methods and MS method) give a very efficient IIR filter design compared to optimal Hankel methods. Actually, when r = 12, suboptimal approximations give better designs compared to optimal Hankel methods. This is clearly seen when comparing the magnitude response and impulse response of both methods. Actually, this indicates that although suboptimal Hankel methods and MS method are not proved to be optimal, they give a very efficient IIR filter designs which may compare favorably to the IIR filters designed using optimal Hankel methods.

TABLE 13

Impulse Response and Singular Values of Example 5.

Impulse response		Singular values	
0.00102126	0.04516681	1.00207139	0.00086072
0.00171127	- 0.24200831	1.00179469	0.00084372
- 0.00117535	- 0.09232160	0.98899439	0.00083630
- 0.00250698	0.11579255	0.98410243	0.00048139
0.00029551	0.05822153	0.90582947	0.00038303
- 0.00000000	- 0.01691367	0.88088190	0.00035131
- 0.00048832	0.01058987	0.68981794	0.00031612
0.00673331	- 0.01139770	0.63389584	0.00028387
0.00494167	- 0.04592731	0.39946810	0.00025306
- 0.01056329	- 0.00000000	0.33663497	0.00022613
- 0.00846484	0.03197690	0.17649909	0.00020267
0.00375803	0.00549619	0.13493693	0.00018348
- 0.00349791	- 0.00349791	0.06325632	0.00016901
0.00549619	0.00375803	0.04429768	0.00015993
0.03197690	- 0.00846484	0.02032435	0.00004155
- 0.00000000	- 0.01056330	0.01344986	0.00003570
- 0.04592731	0.00494167	0.00650714	0.00002608
- 0.01139770	0.00673331	0.00418759	0.00002233
0.01058987	- 0.00048832	0.00225548	0.00000398
- 0.01691367	- 0.00000000	0.00156092	0.00000272
0.05822153	0.00029551	0.00103198	0.00000116
0.11579255	- 0.00250698	0.00098906	0.00000086
- 0.09232160	- 0.00117535	0.00091579	0.00000019
- 0.24200831	0.00171127	0.00091232	0.00000009
0.04516681	0.00102126	0.00087325	0.00000000
0.30078733			

TABLE 14

LSE and L_∞ Error Norms of all Methods Applied in Example 5.

		r = 12		r = 16	
		LSE	L_∞	LSE	L_∞
Least squares methods	Shank	0.16613697	1.30898834	0.00053943	2.37273306
	Prony	0.55262757	1.58338543	0.01283823	2.30580378
	Pade	2.38731394×10^5	1.00585559	4.68532174	1.01973339
Suboptimal methods	Kung	0.03785227	0.11174433	0.00400173	0.01146366
	Kimura	0.03783849	0.11092094	0.00386922	0.01137340
	MS	0.03301519	0.09906520	0.00318947	0.01018798
Optimal Hankel methods	OPH	0.04916033	0.08748981	0.00508302	0.00833639
	OPHd	0.04913972	0.08719112	0.00497937	0.00750243
	OPHD	0.05040309	0.08465571	0.00501910	0.00749107
	CF	0.04924688	0.08638022	0.00504817	0.01005565

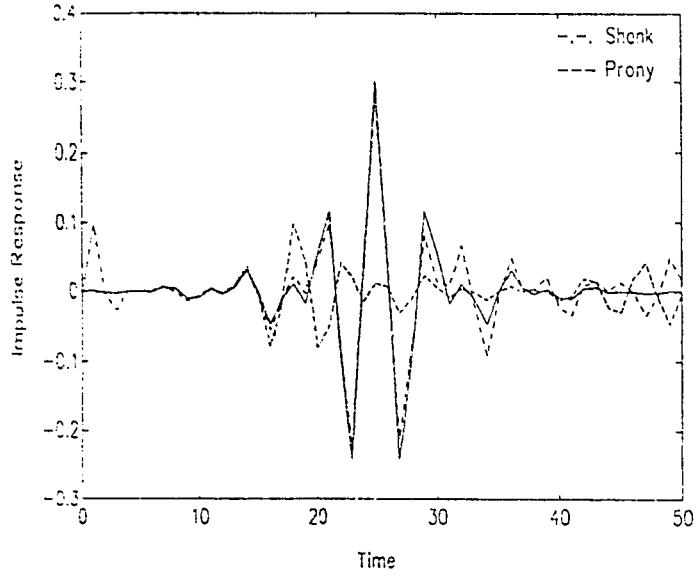


Fig. 5.26a: Impulse response of least squares methods with r = 12.

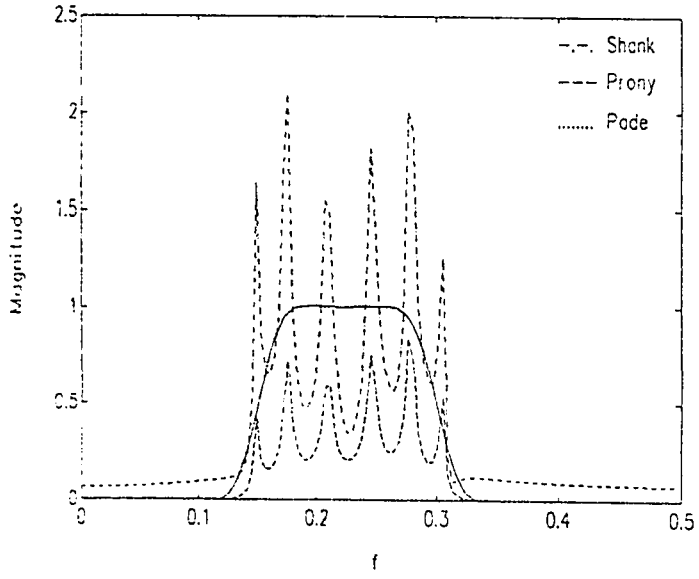


Fig. 5.26b: Magnitude response of least squares methods with r = 12.

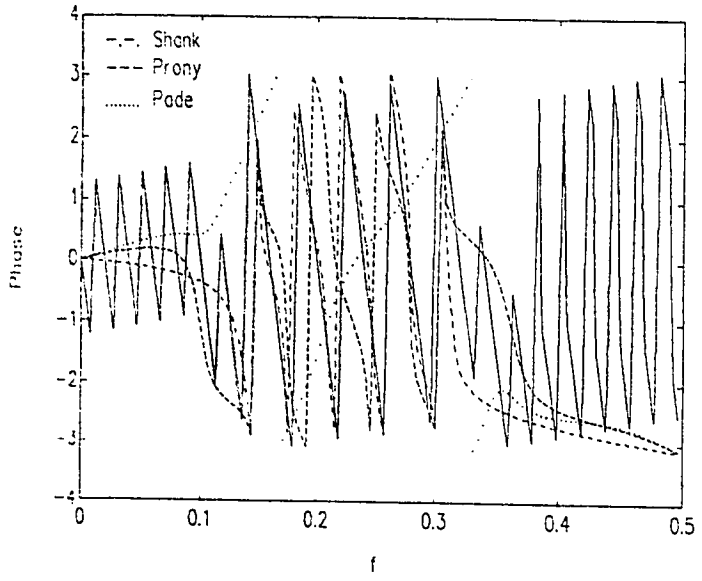


Fig. 5.26c: Phase response of least squares methods with r = 12.

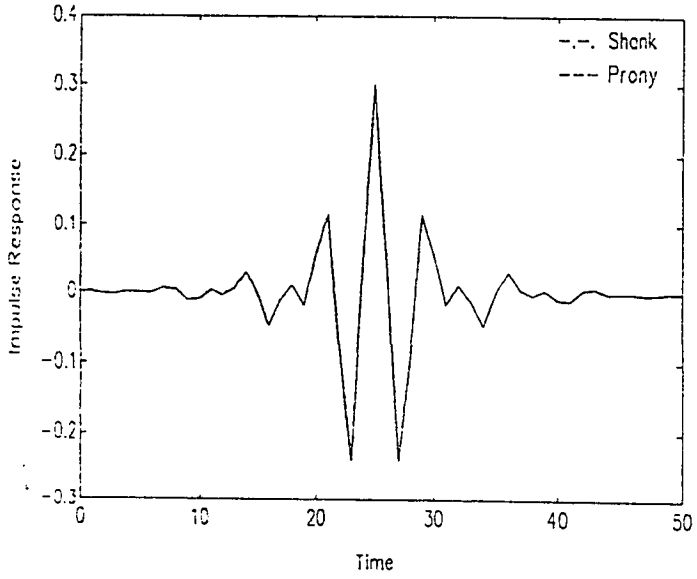


Fig. 5.27a: Impulse response of least squares methods with r = 16.

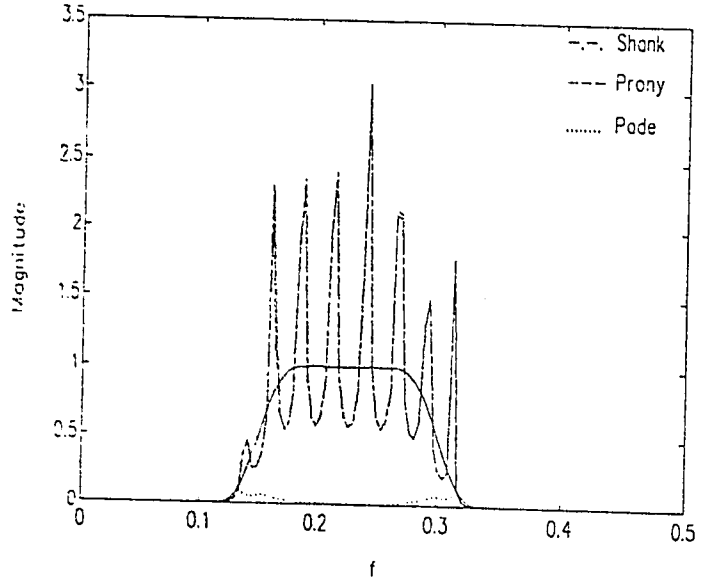


Fig. 5.27b: Magnitude response of least squares methods with r = 16.

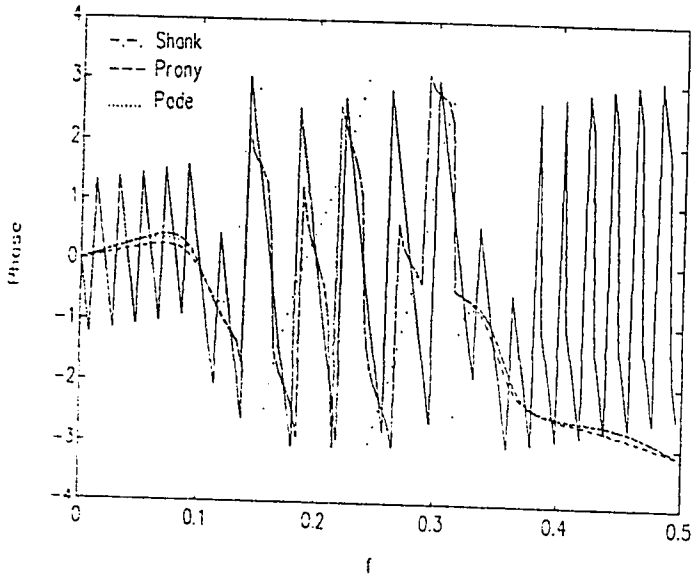


Fig. 5.27c: Phase response of least squares methods with r = 16.

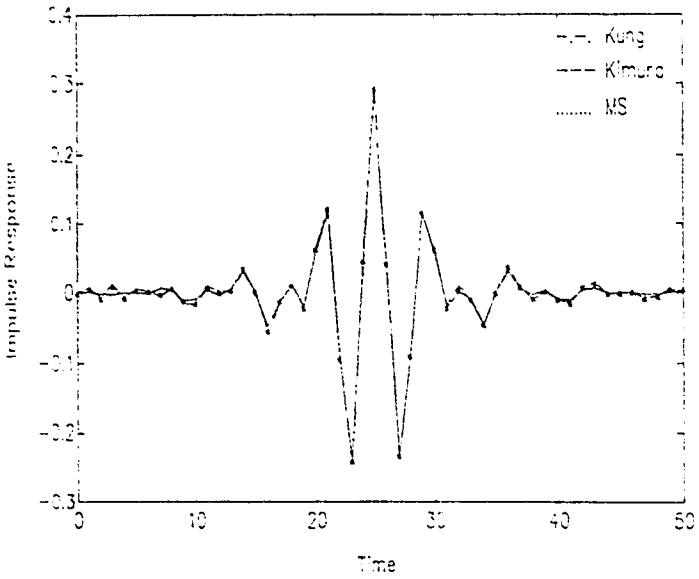


Fig. 5.28a: Impulse response of suboptimal methods with r = 12.

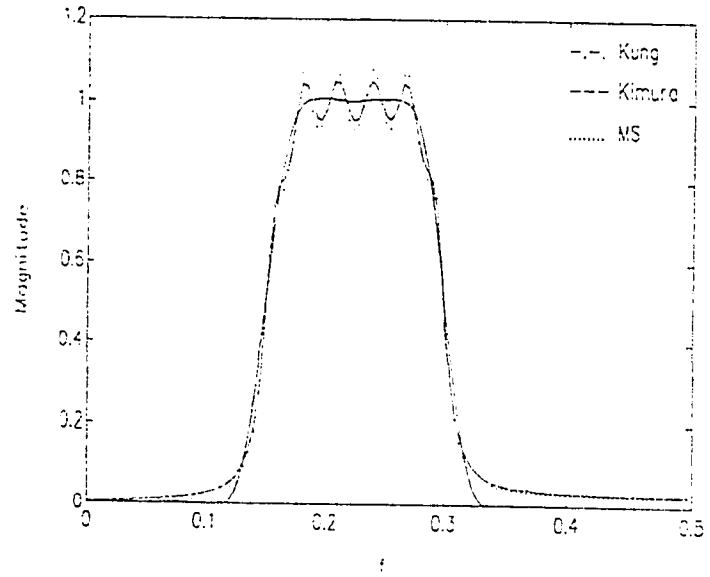


Fig. 5.28b: Magnitude response of suboptimal methods with r = 12.

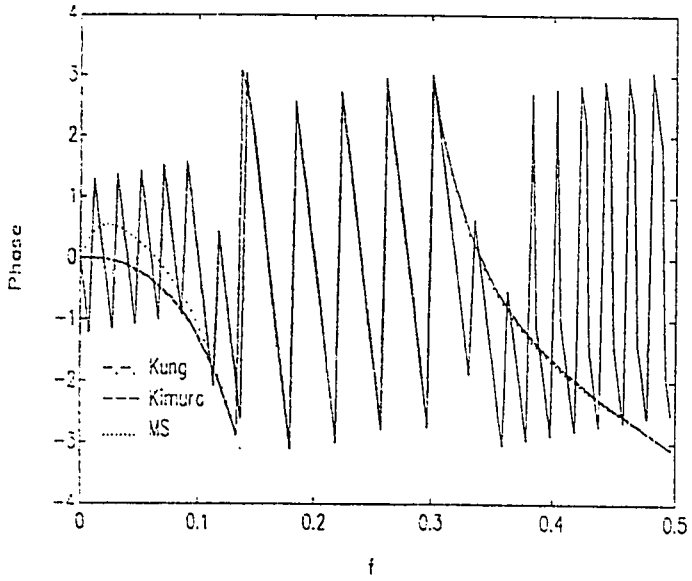


Fig. 5.28c: Phase response of suboptimal methods with r = 12.

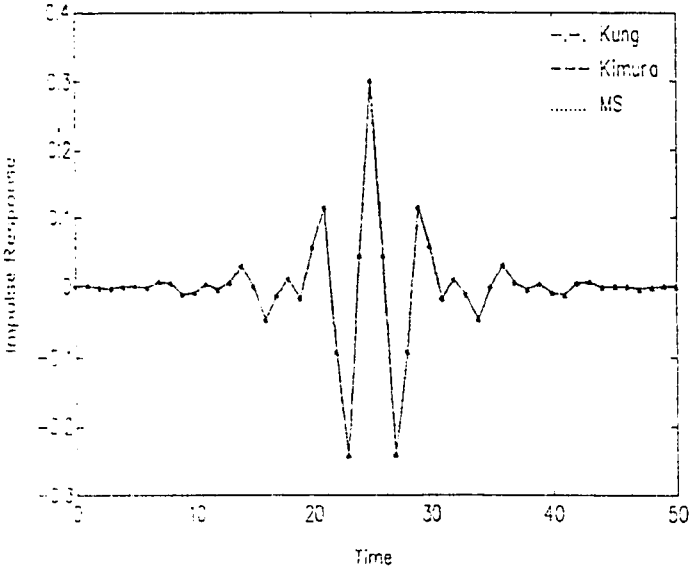


Fig. 5.29a: Impulse response of suboptimal methods with r = 16.

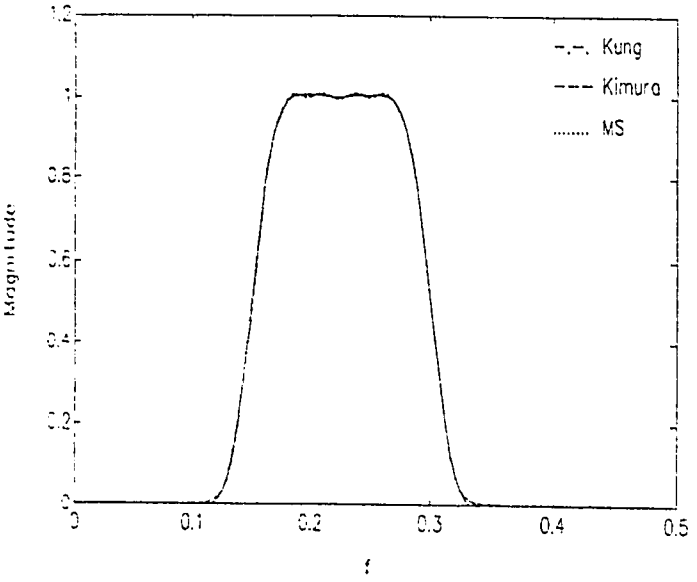


Fig. 5.29b: Magnitude response of suboptimal methods with r = 16.

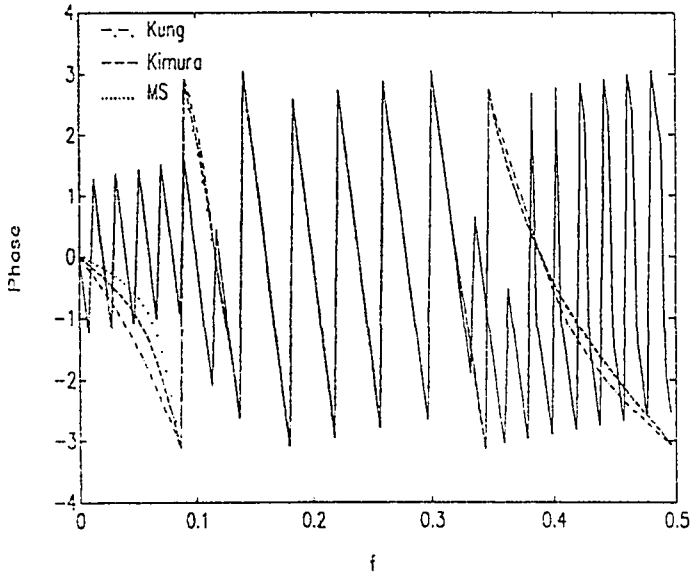


Fig. 5.29c: Phase response of suboptimal methods with r = 16.

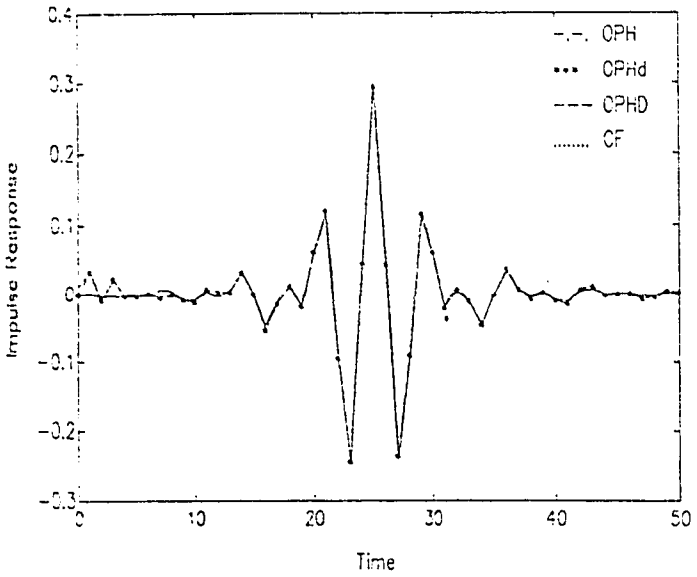


Fig. 5.30a: Impulse response of optimal Hankel methods with r = 12.

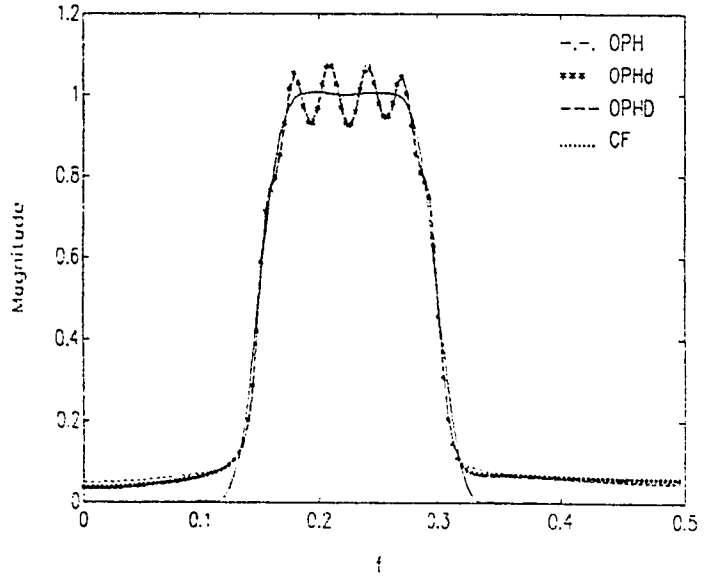


Fig. 5.30b: Magnitude response of optimal Hankel methods with r = 12.

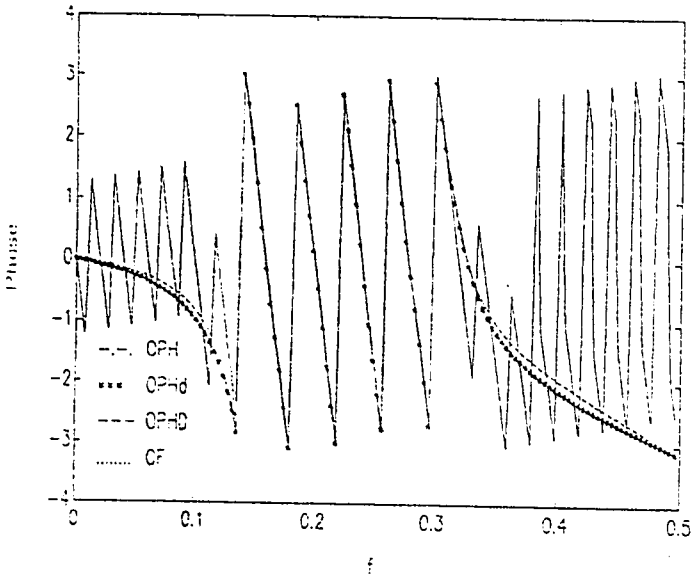


Fig. 5.30c: Phase response of optimal Hankel methods with r = 12.

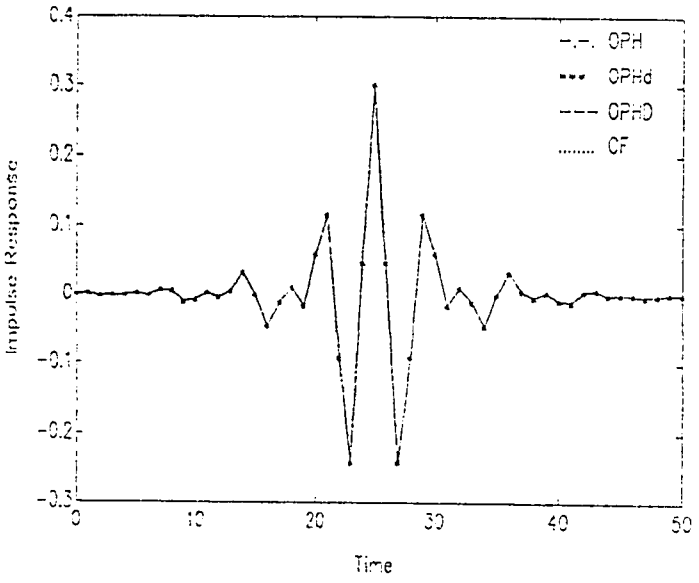


Fig. 5.31a: Impulse response of optimal Hankel methods with r = 16.

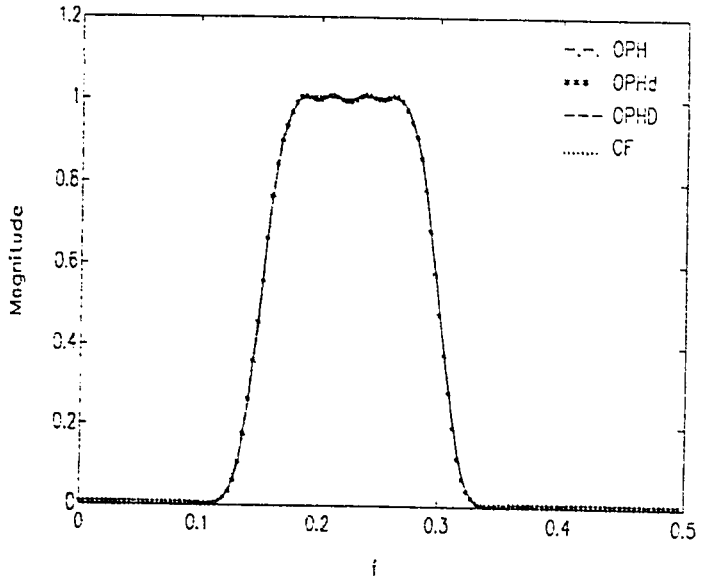


Fig. 5.31b: Magnitude response of optimal Hankel methods with r = 16.

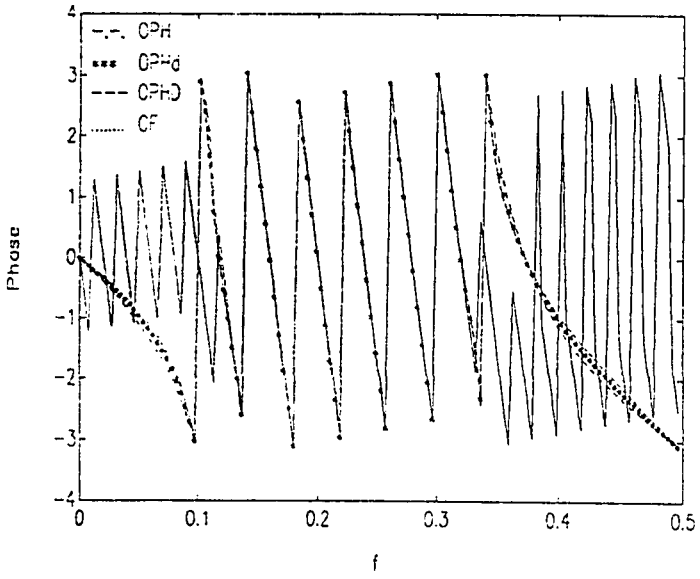


Fig. 5.31c: Phase response of optimal Hankel methods with r =16.

5.2.6 Example 6: Band-Reject Inverse Chebyshev Filter

A 14-th order IIR band-reject filter is designed using the inverse Chebyshev method with f_{c1}= 0.1 and f_{c2}= 0.25. The coefficients of this filter are listed in table 15. This filter is approximated by a (10,10) IIR filter using the first 41 samples of the impulse response given in table 16.

The performance of least squares methods is shown in Fig.'s 5.32(a,b,c) for impulse response, magnitude response, and phase response respectively. The performance of Shank method and Prony method is almost identical and comparable to the performance of Pade approximation in this example. For the impulse response, the three methods give a very good approximations but Pade approximation is less efficient after n = 22 compared to Shank and Prony. For the magnitude response, the performance of the three methods is excellent except at f_{c1} where it deviates specially for Pade approximation. For the phase response, these methods give an almost exact approximation at the pass-bands but this not the case for the stop-band where the approximation is poor. Actually, least squares methods showed a better performance in this example compared to suboptimal methods and optimal Hankel methods.

For suboptimal methods, the impulse response is shown in Fig. 5.33a, the magnitude response in Fig. 5.33b, and the phase response is shown in Fig.5.33c. For the impulse response, the three methods give a very efficient approximations which are identical to each other except at n = 0 where Kung method starts with 0. The magnitude responses of Kimura and MS methods are considered to be superior compared to the magnitude response of Kung which shows a poor performance. For the phase response, Kimura and MS

method give an exact approximations at the pass-bands but for the stop-band, the approximation is poor. Kung method gives a good approximation at the pass-bands but this is not the case at the stop-band.

For optimal Hankel methods, the impulse responses of the four methods are almost exact and identical except at n = 0 where OPH method is equal to 0 as shown in Fig. 5.34a. For the magnitude response, OPH method gives a poor approximation as shown in Fig. 5.34b. OPHd and CF methods give an identical and on the same time good approximations to the magnitude response. OPHD shows a similar performance but it shows a slightly better performance at low frequencies. However, OPHd and CF methods are slightly better at high frequencies. For the phase response, OPHd, OPHD, and CF methods have exact and identical responses at the pass-bands. However, this is not the case for the stop-band where the approximations are poor and OPHD method deviates from OPHd and CF methods. This is shown in Fig. 5.34c.

Finally, this example is concluded by the following observations. Firstly, although least squares methods are not optimal, they could give a better designs compared to optimal Hankel methods specially for non-symmetric impulse response. Secondly, the improvement of OPH method is great when the D-term is added to it. Thirdly, Kimura method gives a much better approximation to the magnitude response at the pass-bands compared to optimal Hankel methods, however the converse is true at the stop-band.

TABLE 15

Coefficients of the Band-Reject Chebyshev Filter

b	a
0.2685	1.0000
-1.7065	-5.2286
6.2026	15.2448
-15.7648	-31.2534
30.9555	49.8731
-48.9097	-64.5876
63.8848	69.6000
-69.7061	-63.0148
63.8848	48.1376
-48.9097	-30.8338
30.9555	16.3720
-15.7648	-7.0104
6.2026	2.3234
-1.7065	-0.5395
0.2685	0.0721

TABLE 16

Impulse Response and Singular Values of Example 6

Impulse response	Impulse response	Singular values	Singular values
0.26849405	0.01270617	0.88894095	0.00056498
-0.30267048	-0.01981608	0.88008746	0.00052326
0.52696167	-0.01124310	0.60489861	0.00040072
-0.00397174	-0.00373870	0.57165649	0.00032365
0.05121425	-0.02444040	0.27994532	0.00032313
0.32447458	-0.02928527	0.24051737	0.00011539
0.15928949	-0.00073114	0.10487067	0.00010821
-0.00095068	0.01671615	0.09446633	0.00000033
0.10779539	0.00834355	0.06065904	
0.10331149		0.05886221	
-0.07423800		0.04452264	
-0.10253092		0.04233393	
-0.01107543		0.03827770	
-0.04098492		0.03665873	
-0.10504927		0.03224811	
-0.03038262		0.03084439	
0.05829912		0.02624538	
0.02848184		0.02532786	
0.00138062		0.01890201	
0.05565550		0.01692269	
0.06794247		0.01614140	
-0.00236643		0.01526524	
-0.03409136		0.00722802	
-0.00634513		0.00710536	
-0.01744161		0.00511829	
-0.05612902		0.00328589	
-0.03313068		0.00277178	
0.02071748		0.00149199	
0.02114270		0.00145617	
0.00023013		0.00073630	
0.02143989		0.00063915	
0.04249901		0.00063443	

TABLE 17

LSE and L_∞ Error Norms of all Methods Applied in Example 6.

		r = 10	
		LSE	L_∞
Least squares methods	Shank	0.01249284	0.19127590
	Prony	0.01265492	0.18959128
	Pade	0.03674833	0.37034505
Suboptimal methods	Kung	0.26974643	0.44823802
	Kimura	0.02596308	0.25738302
	MS	0.03340232	0.09774863
Optimal Hankel methods	OPH	0.27086743	0.47664941
	OPHd	0.03577869	0.23257457
	OPHD	0.03795418	0.24293269
	CF	0.03635783	0.24538318

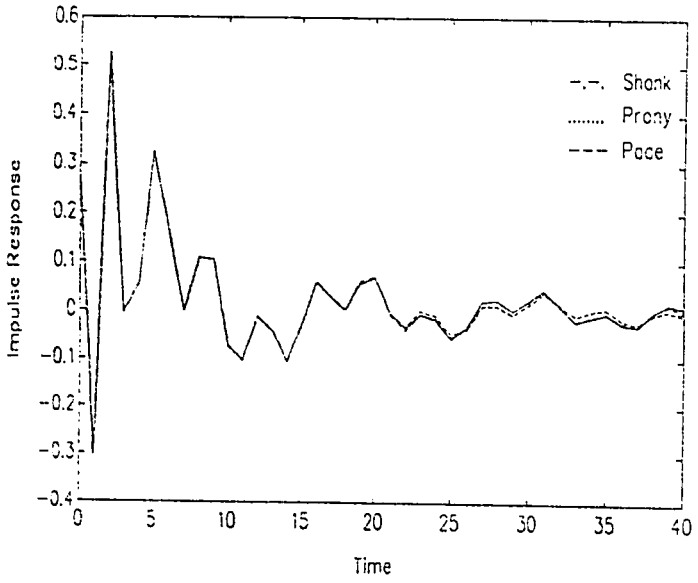


Fig. 5.32a: Impulse response of least squares methods with r = 10.

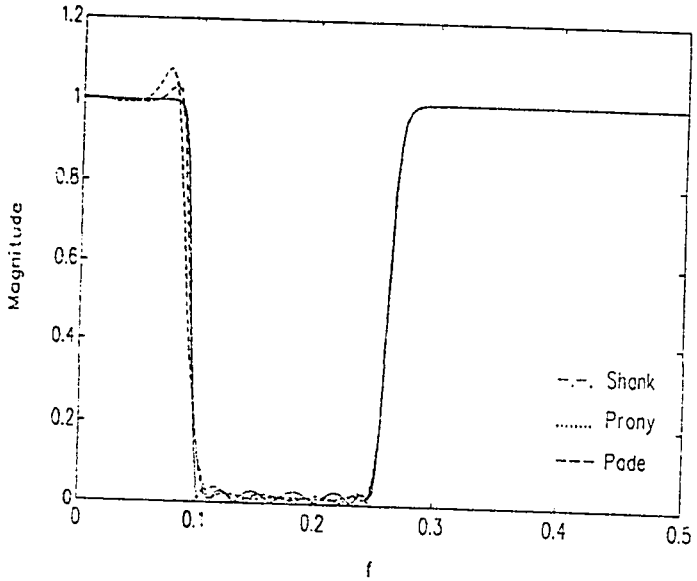


Fig. 5.32b: Magnitude response of least squares methods with r = 10.

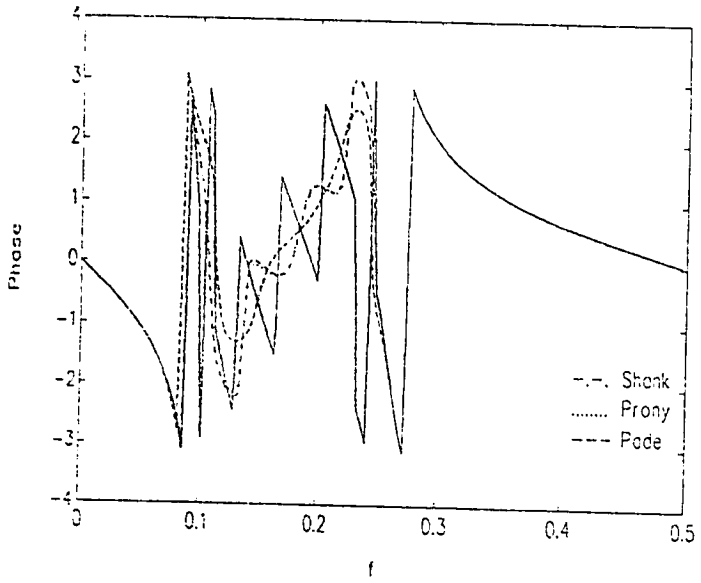


Fig. 5.32c: Phase response of least squares methods with r = 10.

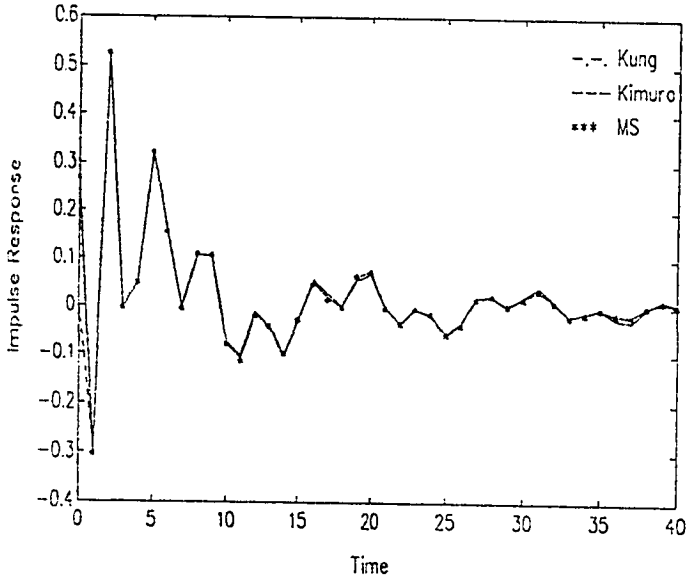


Fig. 5.33a: Impulse response of suboptimal methods with r = 10.

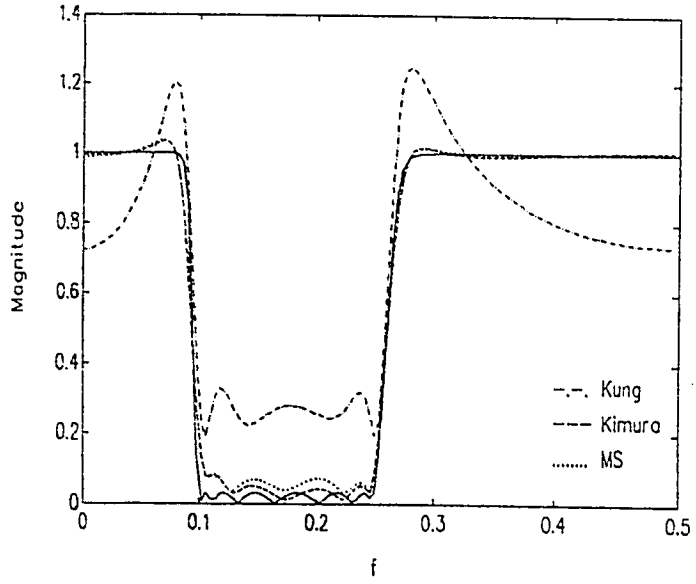


Fig. 5.33b: Magnitude response of suboptimal methods with r = 10.

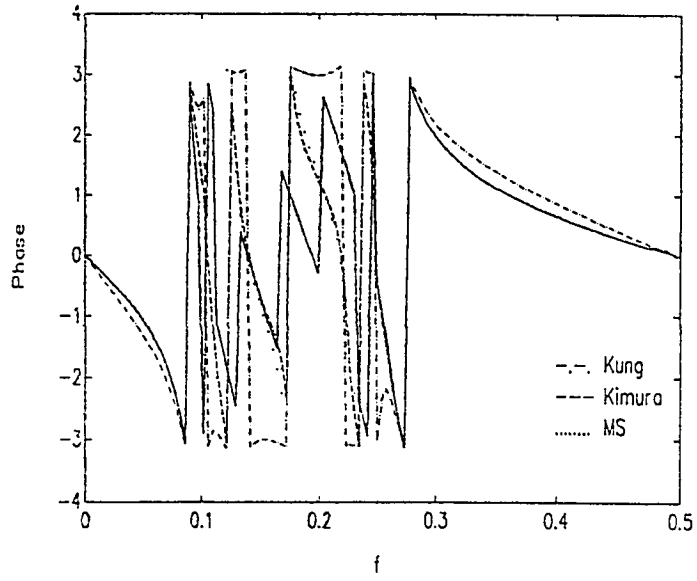


Fig. 5.33c: Phase response of suboptimal methods with r = 10.

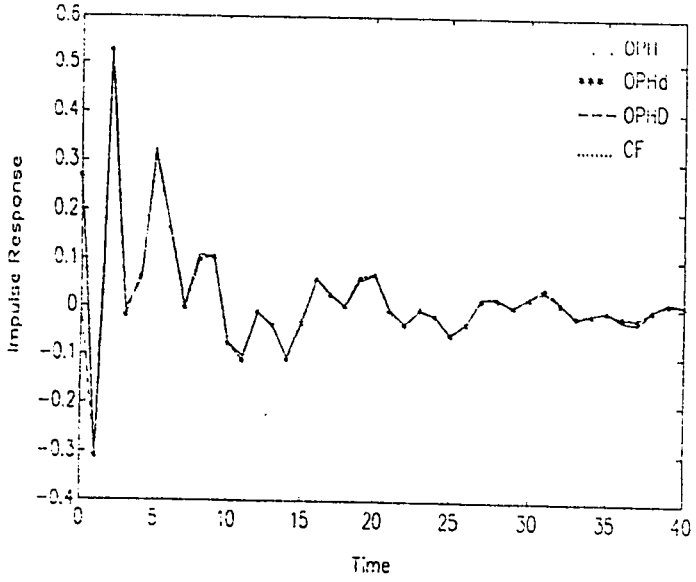


Fig. 5.34a: Impulse response of optimal Hankel methods with r = 10.

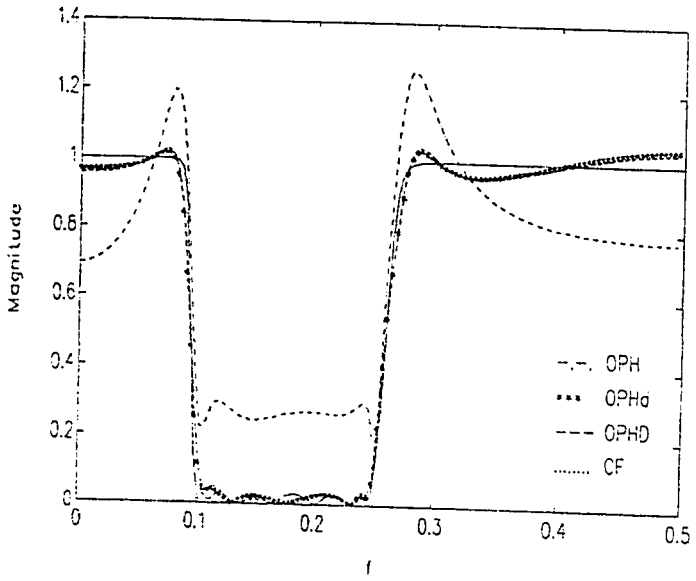


Fig. 5.34b: Magnitude response of optimal Hankel methods with r = 10.

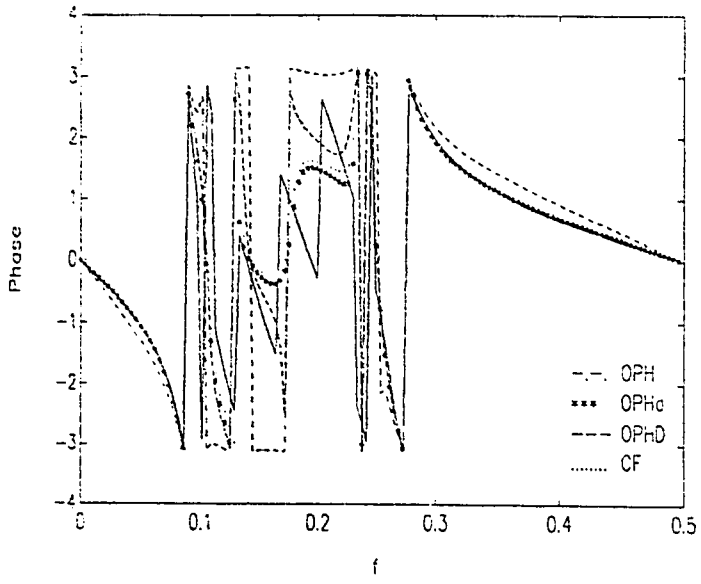


Fig. 5.34c: Phase response of optimal Hankel methods with r = 10.

5.3 Two-Sided Approximation Technique

In this section, some examples which show the efficiency of two-sided approximation technique in approximating the magnitude response are provided. The examples discussed in the previous section will be repeated here but with a lower order IIR filters. The performance of two-sided approximation and one-sided approximation (the regular way of approximation where the impulse response is truncated and shifted to obtain a finite and causal data) techniques are compared. The error bound given in theorem 4.1 is verified for each example and shown in a tabulated form.

The concentration during discussion will be on the magnitude response since , as mentioned previously, the impulse and phase responses are highly affected because of stability and causality requirements. Moreover, the phase responses obtained for different approximation methods are generally similar in each example as will be seen soon. Actually, the phase responses of two-sided approximation technique are totally different from the original phase but they are generally smooth and considered to be marginally linear. The same applies for the impulse responses of two-sided approximation technique where they look like a shifted but distorted version of the original impulse response .

5.3.1 Example 1: LPH LPF

Consider example 2 given in the last section. Two-sided approximation technique is applied to the low pass filter using the design methods discussed

previously. The low pass filter is approximated by a (4,4) IIR filter and the results are compared to the results obtained from one-sided approximation technique in the last section.

The magnitude responses of least squares methods are shown in Fig. 5.35b. These methods give an excellent approximation compared to one-sided approximation shown in Fig. 5.7b. for r = 5 and in Fig. 5.8b for r = 7. The impulse responses are shown in Fig. 5.35a and the phase responses are shown in Fig. 5.35c. Actually, the impulse response seem to be a shifted but distorted version of the original impulse response.

For suboptimal methods, Kimura method and MS method give a very good results. The magnitude responses of these methods are shown in Fig. 5.36b. Kimura method, when applying two-sided technique to it, gives a better approximation to the desired magnitude response compared to the same method when one-sided technique is applied to it to obtain a (5,5) IIR filter (Fig. 5.9b). The same applies to MS method. The magnitude response of Kung method has a large approximation error. The impulse responses are given in Fig. 5.36a. Note that the impulse responses of Kimura and MS methods are similar to what we had in least squares methods. The phase responses of these methods are shown in Fig. 5.36c.

The magnitude responses of optimal Hankel methods are shown in Fig. 5.37b. OPHd, OPHD, and CF methods give a better approximation specially at the stop-band when compared to the magnitude response of one-sided approximation shown in Fig. 5.11b for (5,5) IIR filter. OPH method gives a poor

approximation to the magnitude response. The impulse responses are shown in Fig. 5.37a. and the phase responses are shown in Fig. 5.37c

A question arises is: what is the reason for the bad performance of Kung and OPH methods?. Actually, the answer is the D-term again. The formula given to calculate the coefficients of two-sided approximation is in the form of summation. Thus, the error due to neglecting the D-term will appear in every coefficients of the two-sided approximation designed filter. Note how much the improvement of OPH method performance when the D-term is forced to it which could be seen from the performance of OPHd method.

The error bound given in theorem for the magnitude response is verified and the results are provided in table 18.

TABLE 18

Magnitude Response Error Bound of Example 1.

		$\left\|\hat{H}_a(e^{j\omega})-\hat{H}(e^{j\omega})\right\|$	$\left\|\|H_a(e^{j\omega})\|-\|H(e^{j\omega}\|\right\|$
Least squares methods	Shank	0.03003155	0.06006310
	Prony	0.03016916	0.06033832
	Pade	0.06711034	0.11989247
Suboptimal methods	Kung	0.17155630	0.34311261
	Kimura	0.02336543	0.04292090
	MS	0.06818763	0.04137970
Optimal Hankel methods	OPH	0.17098766	0.33951587
	OPHd	0.02089519	0.03932416
	OPHD	0.02097223	0.03510308
	CF	0.02125565	0.04004522

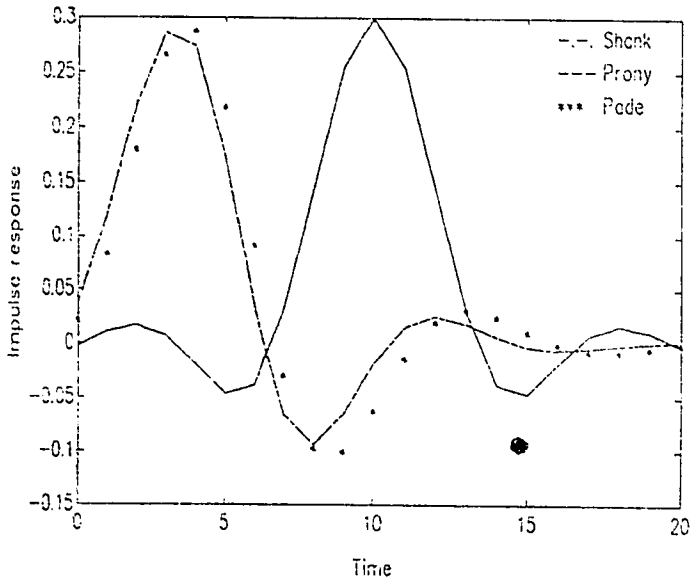


Fig. 5.35a: Impulse response of least squares methods using two-sided approximation technique with r = 4.

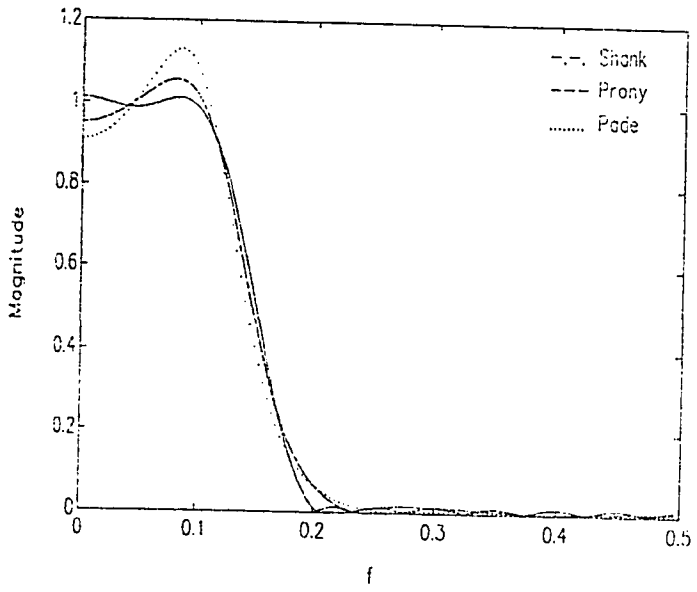


Fig. 5.35b: Magnitude response of least squares methods using two-sided approximation technique with r = 4.

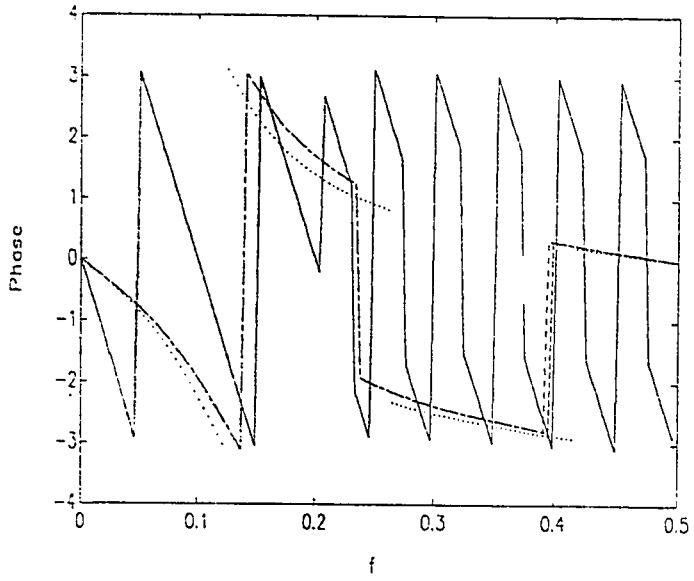


Fig. 5.35c: Phase response of least squares methods using two-sided approximation technique with r = 4.

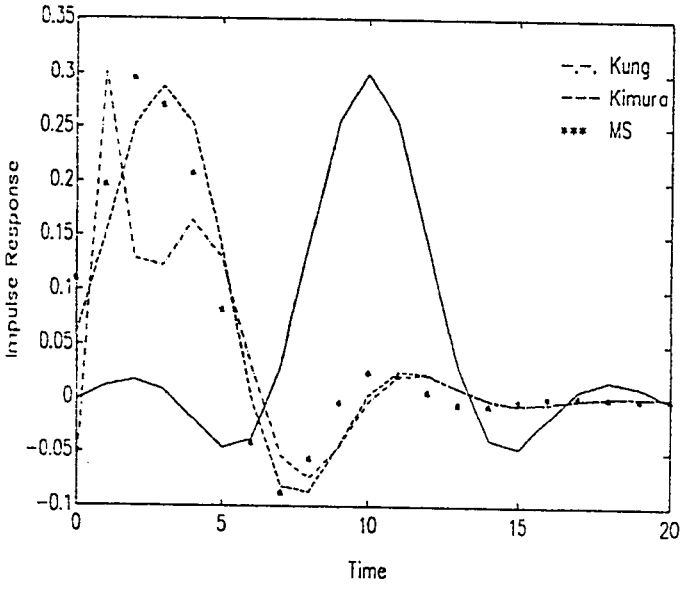


Fig. 5.36a: Impulse response of suboptimal methods using two-sided approximation technique with r = 4.

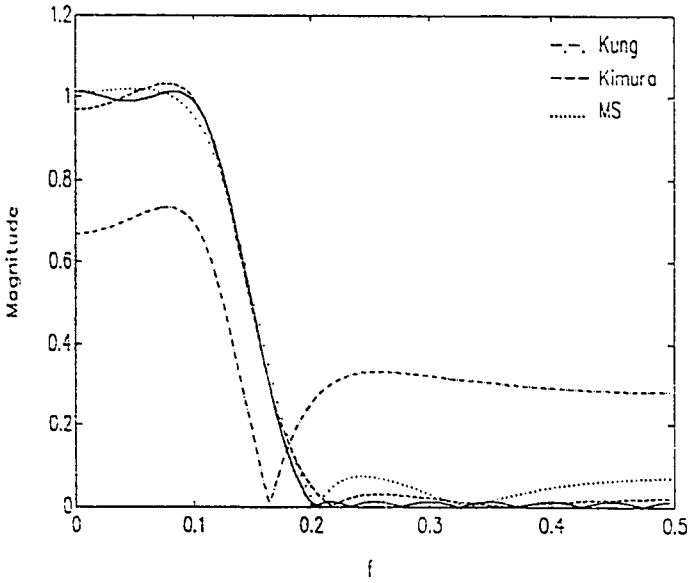


Fig. 5.36b: Magnitude response of suboptimal methods using two-sided approximation technique with r = 4.

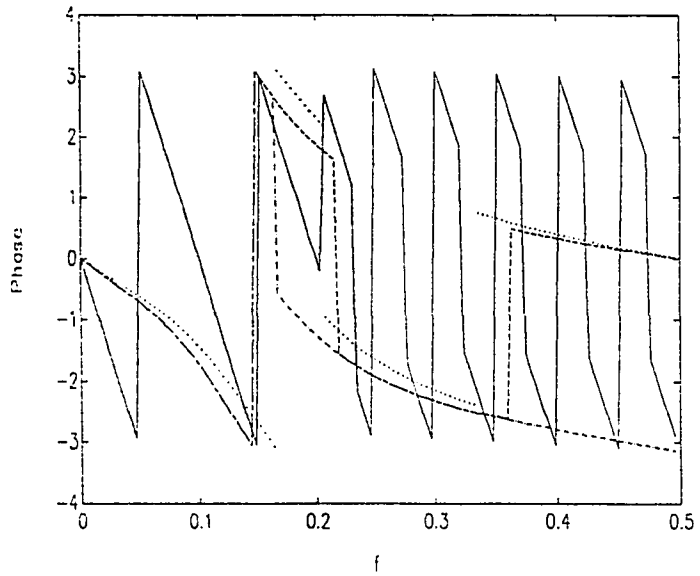


Fig. 5.36c: Phase response of suboptimal methods using two-sided approximation technique with r = 4.

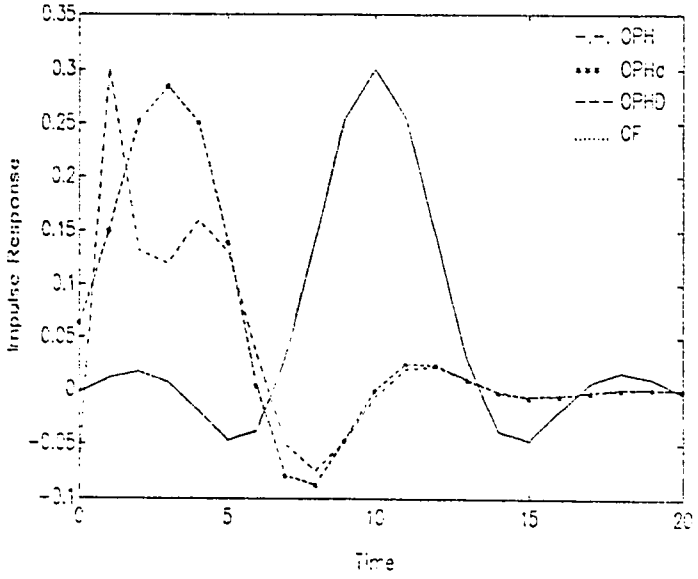


Fig. 5.37a: Impulse response of optimal Hankel methods using two-sided approximation technique with r = 4.

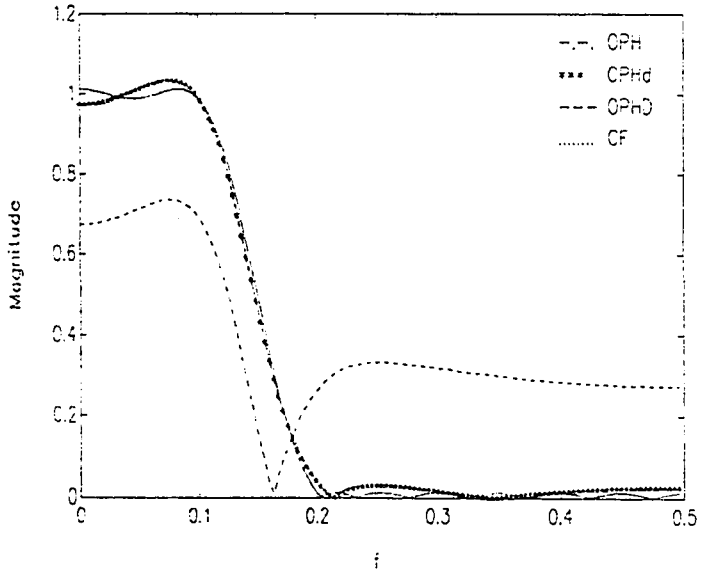
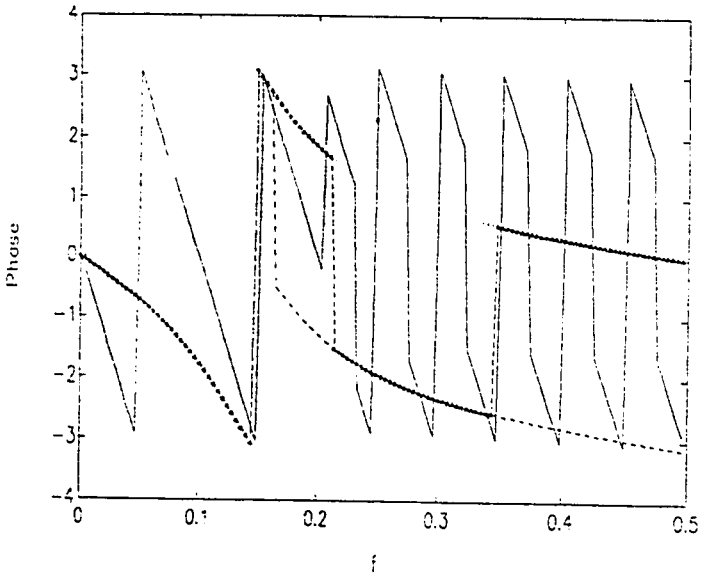


Fig. 5.37b: Magnitude response of optimal Hankel methods using two-sided approximation technique with r = 4.



5.37c: Phase response of optimal Hankel methods using two-sided approximation technique with r = 4.

5.3.2 Example 2: Ideal Differentiator

Example 3 is repeated here where a (4,4) IIR filter is designed by applying two-sided approximation to approximate an ideal differentiator.

The impulse responses of the implemented methods are generally similar and could be looked at as a distorted shifted version of the original impulse response. This is shown in Fig. 5.38a for least squares, in Fig. 5.39a for suboptimal methods, and in Fig.5.40a for optimal Hankel methods.

The efficiency of two-sided approximation technique is apparent when the magnitude response is considered.

For least squares methods, the magnitude response of the (4,4) IIR filter shown in Fig. 5.38b is superior compared to the magnitude response of the (21,21) IIR filter shown in Fig. 5.15b designed via one-sided approximation technique. The phase response is shown in Fig. 5.38c. Shank and Prony methods have an identical phase responses which are also similar to Pade's. The Phase responses of least squares methods don't approximate the original phase response by any way but they are marginally linear and smooth.

For suboptimal methods, the magnitude responses shown in Fig 5.39b are excellent compared to the magnitude responses of the (21,21) filter designed using one-sided approximation technique and shown in Fig. 5.16b. The phase responses of these methods are almost identical and the same argument applies as least squares methods regarding the linearity and smoothness of these methods. The phase responses are shown in Fig. 5.39c.

The magnitude responses of optimal Hankel methods given in Fig. 5.40b are superior compared to the magnitude responses obtainedusing one-

sided approximation technique which are shown in Fig. 5.17b. The Phase responses are shown in Fig. 5.40c and they are almost identical to the Phase responses of least squares methods and suboptimal methods.

TABLE 19

Magnitude Response Error Bound of Example 2.

		$\left\|\hat{H}_a(e^{j\omega})-\hat{H}(e^{j\omega})\right\|$	$\left\|\|H_a(e^{j\omega})\|-\|H(e^{j\omega}\|\right\|$
Least squares methods	Shank	0.08110264	0.15719869
	Prony	0.08102288	0.15595009
	Pade	0.31065217	0.53619016
Suboptimal methods	Kung	0.02466598	0.04837185
	Kimura	0.02466598	0.04837185
	MS	0.03173199	0.06285290
Optimal Hankel methods	OPH	0.02220070	0.04051634
	OPHd	0.02220070	0.04051634
	OPHD	0.02164283	0.04051634
	CF	0.03005454	0.04051634

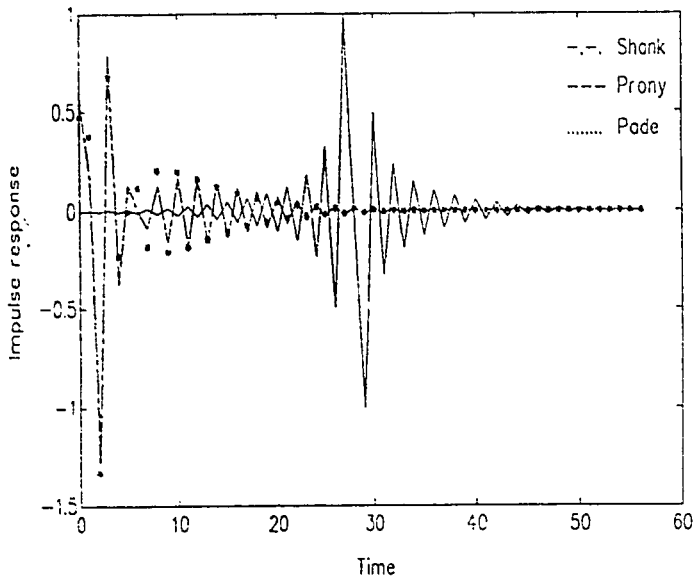


Fig. 5.38a: Impulse response of least squares methods using two-sided approximation technique with r = 4.

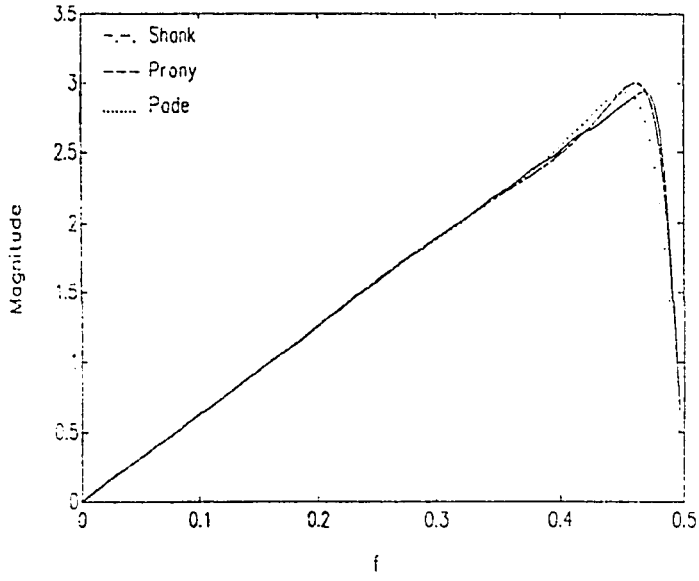


Fig. 5.38b: Magnitude response of least squares methods using two-sided approximation technique with r = 4.

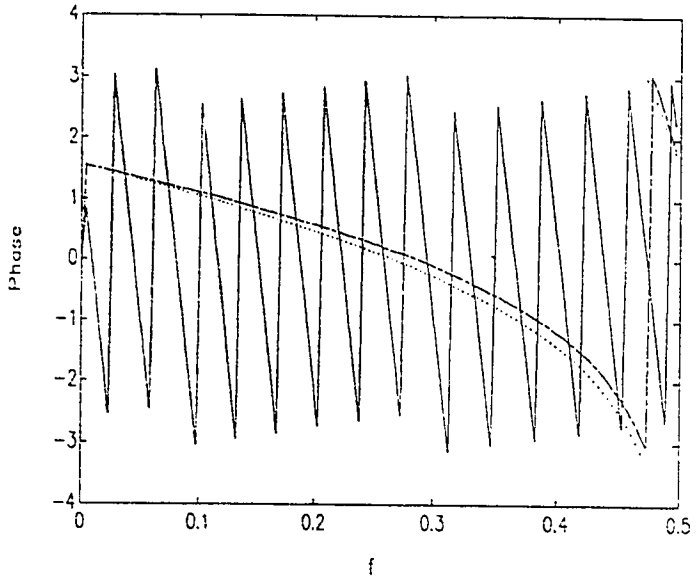


Fig. 5.38c: Phase response of least squares methods using two-sided approximation technique with r = 4.

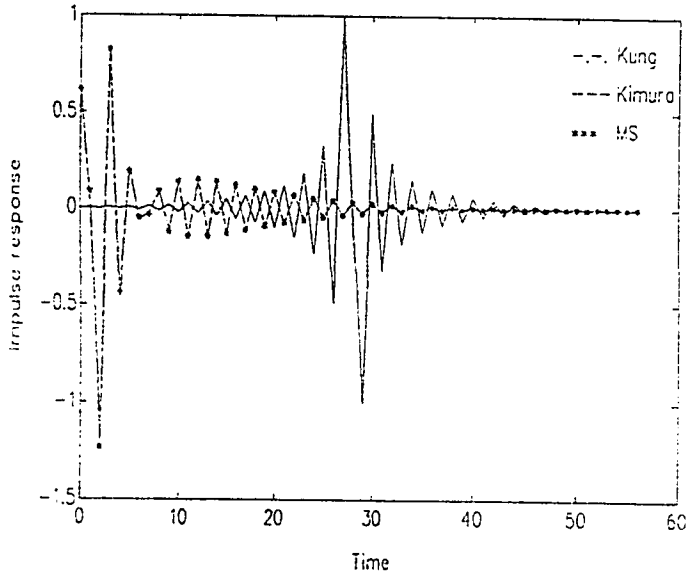


Fig. 5.39a: Impulse response of suboptimal methods using two-sided approximation technique with r = 4.

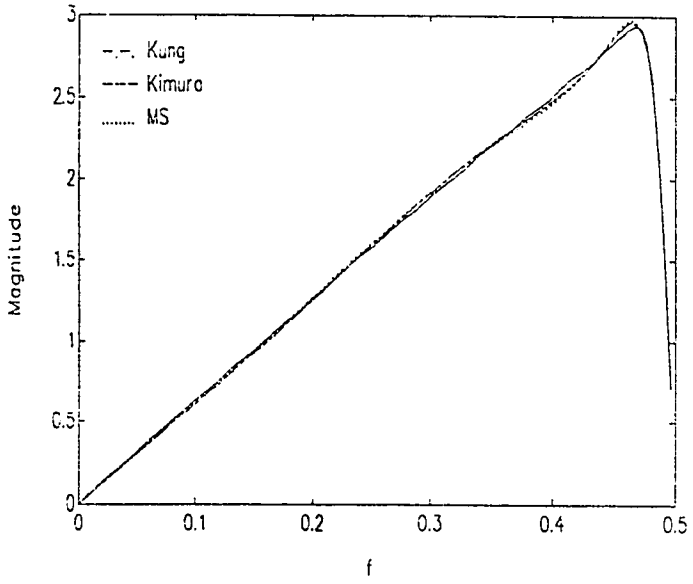


Fig. 5.39b: Magnitude response of suboptimal methods using two-sided approximation technique with r = 4.

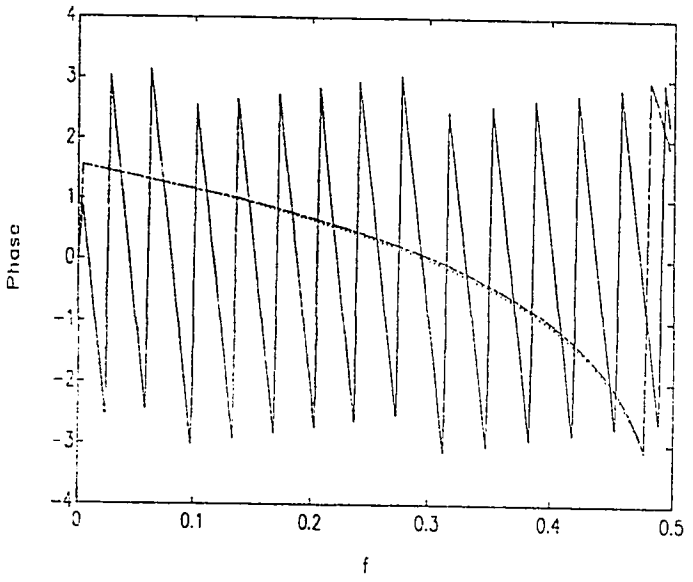


Fig. 5.39c: Phase response of suboptimal methods using two-sided approximation technique with r = 4.

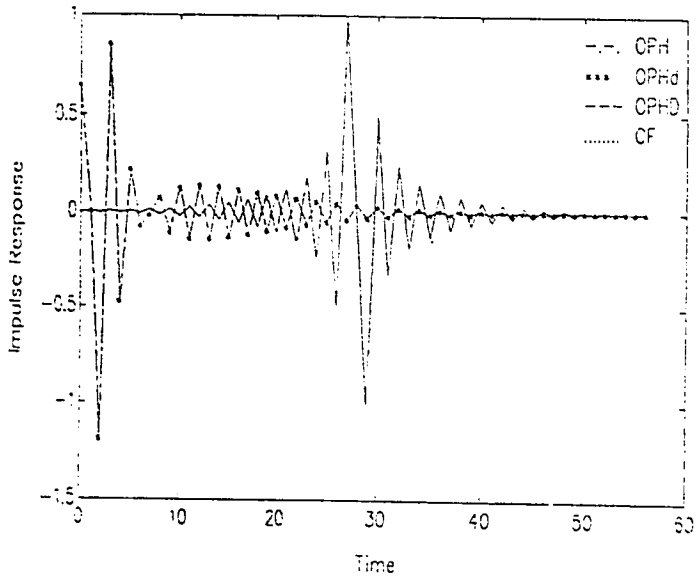


Fig. 5.40a: Impulse response of optimal Hankel methods using two-sided approximation technique with r = 4.

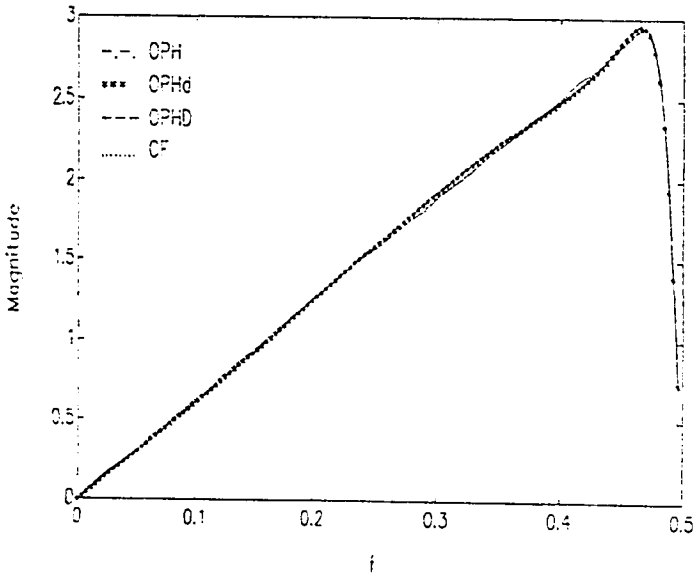


Fig. 5.40b: Magnitude response of optimal Hankel methods using two-sided approximation technique with r = 4.

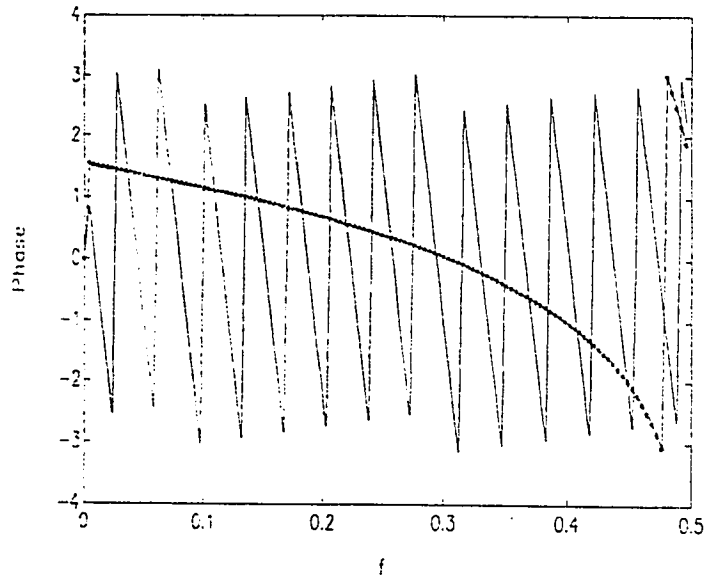


Fig. 5.40c: Phase response of optimal Hankel methods using two-sided approximation technique with r = 4.

5.3.3 Example 3: Ideal BPF

Based on the impulse response given in example 4, a (6,6) IIR filter is designed to approximate an ideal HPF. The designed filter is compared to the (9,9) and (11,11) IIR filters designed by applying one-sided approximation technique in 5.2.4 example 4.

The impulse responses of least squares methods, suboptimal methods, and optimal Hankel methods are shown in Fig. 5.41a, Fig. 5.42a, and Fig. 5.43a respectively. The impulse responses of these methods are generally similar and they seem to be a shifted but distorted version of the original impulse response.

The magnitude responses of least squares methods shown in Fig. 5.41b are greatly improved compared to the magnitude response of the IIR filter designed using one-sided approximation technique with r = 11 (Fig. 5.21b). The Phase responses of these methods are comparable and shown in Fig. 5.41c.

For suboptimal methods, Kimura method and MS method give an excellent approximation to the magnitude response as shown in Fig. 5.42b. Moreover, the approximation is comparable to the magnitude response of the (11,11) IIR filter designed using one-sided approximation and shown in Fig. 5.23b. Kung method gives a poor approximation to the magnitude response due to the D-term effect. The phase responses of these methods are generally similar and shown in Fig. 5.42c.

For optimal Hankel methods, the magnitude responses are shown in Fig. 5.43b. OPH method gives a poor approximation. However, the approximation extremely improved when OPHd method is applied. Actually, OPHd, OPHD,

and CF methods have an almost identical performance which is comparable to the magnitude response of (11,11) IIR filter designed using one-sided approximation technique (Fig. 5.25b). The phase responses are generally similar and shown in Fig. 5.43c.

TABLE 20

Magnitude Response Error Bound of Example 3.

		$\left\|\hat{H}_a(e^{j\omega})-\hat{H}(e^{j\omega})\right\|$	$\left\|\|H_a(e^{j\omega})\|-\|H(e^{j\omega}\|\right\|$
Least squares methods	Shank	0.01353189	0.02474063
	Prony	0.01348197	0.02451833
	Pade	0.05078856	0.08845932
Suboptimal methods	Kung	0.15531248	0.31061166
	Kimura	0.00644465	0.01150193
	MS	0.01727353	0.00884858
Optimal Hankel methods	OPH	0.15413976	0.30721872
	OPHd	0.00487141	0.00911335
	OPHD	0.00482286	0.00884212
	CF	0.00547049	0.01054463

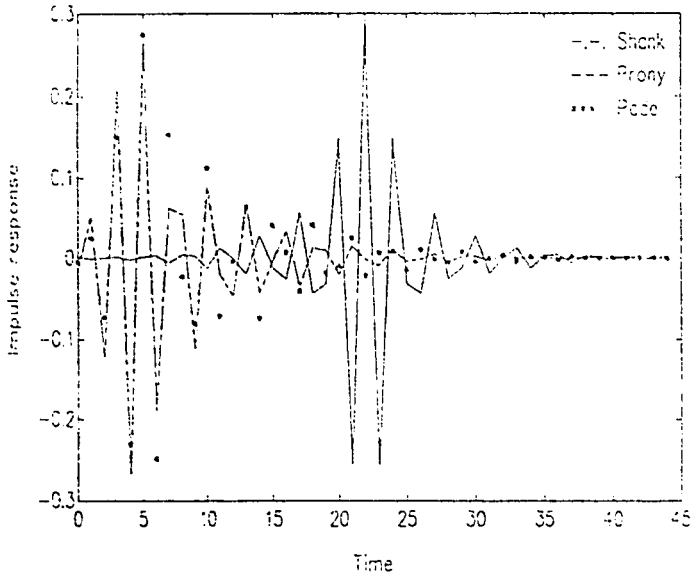


Fig. 5.41a: Impulse response of least squares methods using two-sided approximation technique with r = 6.

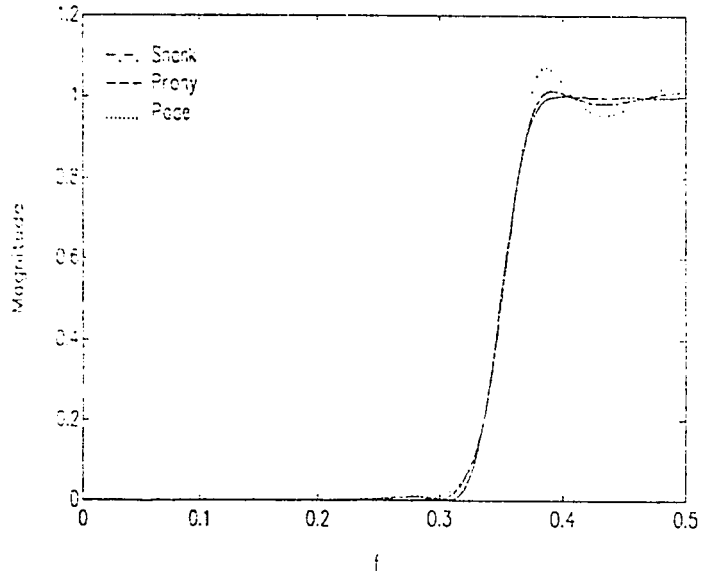


Fig. 5.41b: Magnitude response of least squares methods using two-sided approximation technique with r = 6.

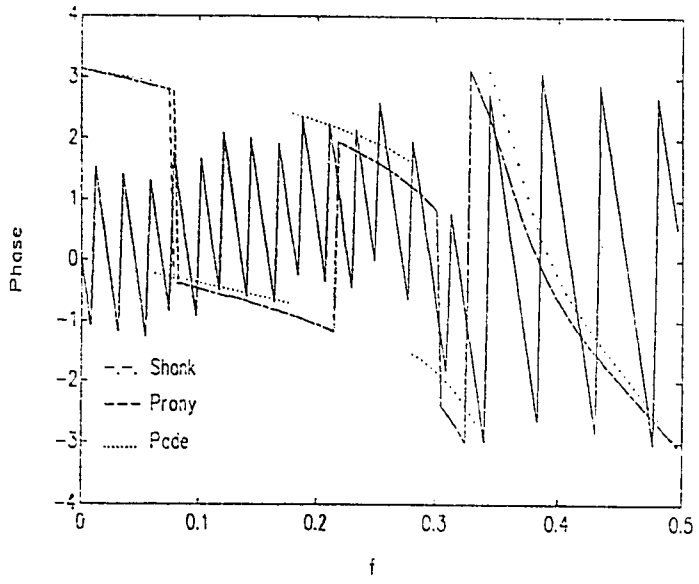


Fig. 5.41c: Phase response of least squares methods using two-sided approximation technique with r = 6.

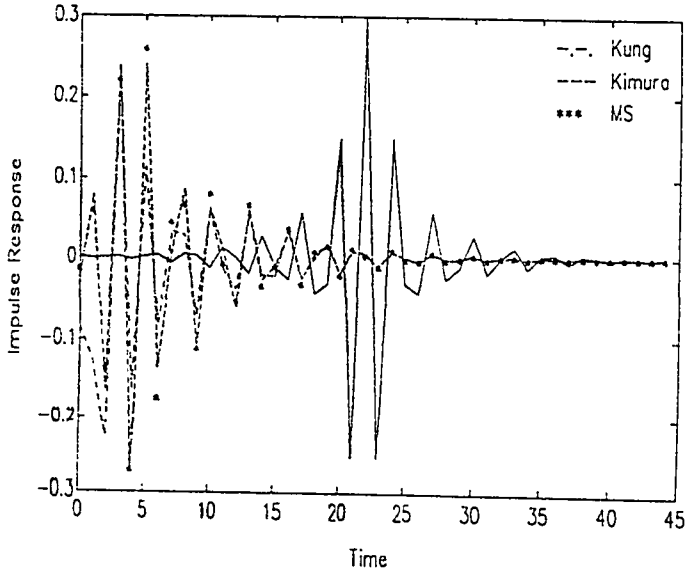


Fig. 5.42a: Impulse response of suboptimal methods using two-sided approximation technique with r = 6.

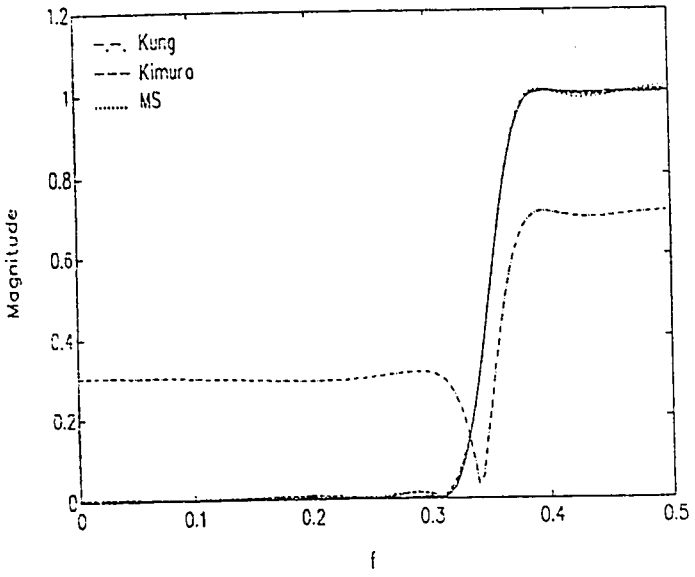


Fig. 5.42b: Magnitude response of suboptimal methods using two-sided approximation technique with r = 6.

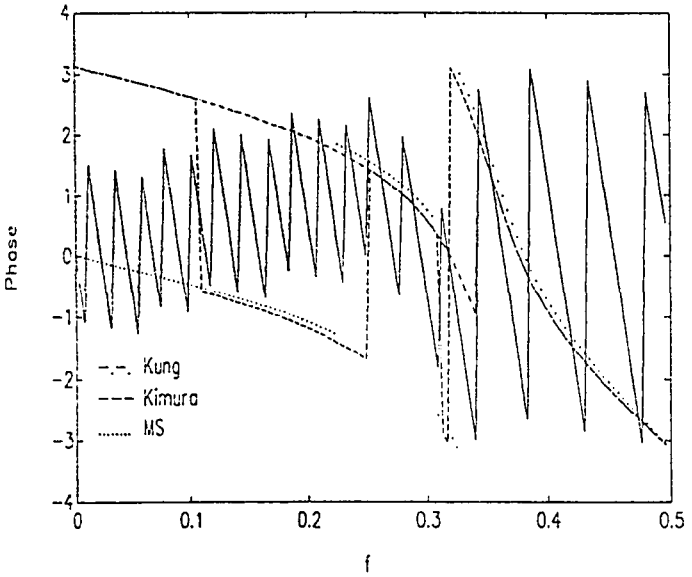


Fig. 5.42c: Phase response of suboptimal methods using two-sided approximation technique with r = 6.

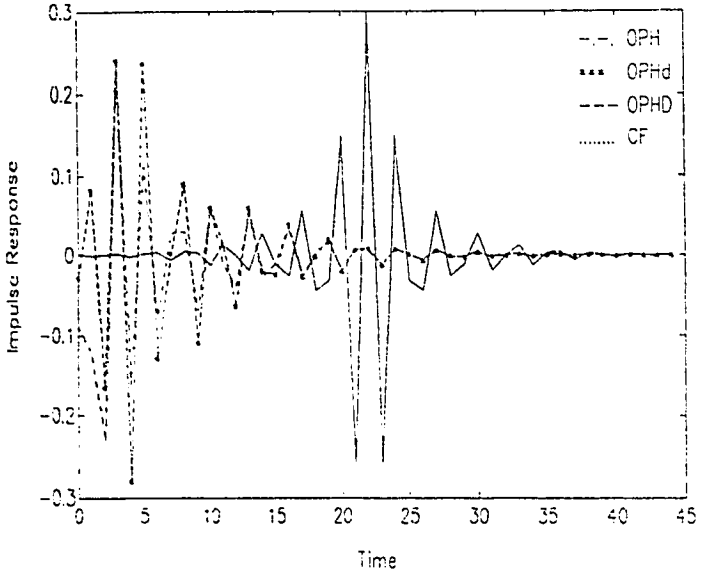


Fig. 5.43a: Impulse response of optimal Hankel methods using two-sided approximation technique with r = 6.

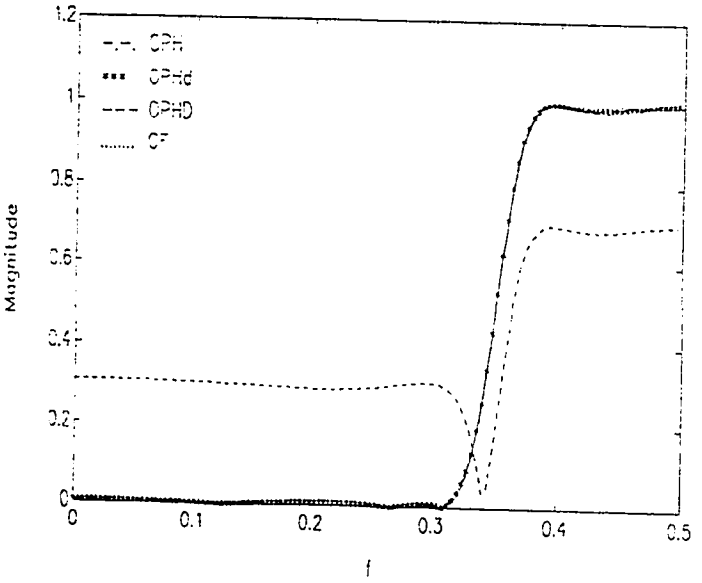


Fig. 5.43b: Magnitude response of optimal Hankel methods using two-sided approximation technique with r = 6.

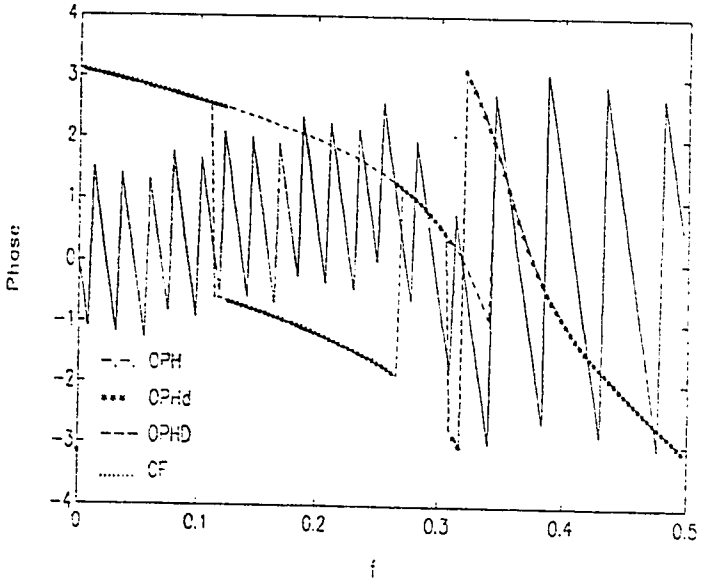


Fig. 5.43c: Phase response of optimal Hankel methods using two-sided approximation technique with r = 6.

5.3.4 Example 4: Ideal HPF

The ideal BPF given in example 5 is approximated with a (8,8) IIR filter using two-sided approximation technique.

The magnitude responses of least squares methods shown in Fig. 5.44b are greatly improved compared to the magnitude responses of (12,12) and (16,16) IIR filters designed using one-sided approximation and shown in Fig. 5.26b and Fig. 5.27b respectively. The impulse responses are shown in Fig. 5.44c and the phase responses are shown in Fig. 5.44a.

The magnitude response of suboptimal methods are shown in Fig. 5.45b. Kung method gives a poor approximation to the magnitude responseof the BPF. However, Kimura and MS methods show a better performance than the (12,12) IIR filter of one-sided approximation. This could be seen by comparing Fig. 5.45b with Fig. 5.28b. Moreover, the magnitude responses of Kimura method and MS method are comparable to the magnitude responses of the (16,16) IIR filter of one-sided approximation shown in Fig. 5.29b. The impulse responses are shown in Fig. 5.45a and the phase responses are shown in Fig. 5.45c.

The magnitude responses of optimal Hankel methods applied using two-sided approximation are shown in Fig. 5.46b. OPHd method, OPHD method, and CF methodgive a similar responses which are considered better compared to the magnitude response of the (12,12) IIR filter shown in Fig. 5.30b. OPH method gives a large error in approximating the desired magnitude response. The impulse responses and phase responses of optimal Hankel methods are shown in Fig. 5.46c and Fig. 5.46a respectively.

TABLE 21

Magnitude Response Error Bound of Example 4.

		$\left\|\hat{H}_a(e^{j\omega})-\hat{H}(e^{j\omega})\right\|$	$\left\|\left\|H_a(e^{j\omega})\right\|-\left\|H(e^{j\omega})\right\|\right\|$
Least squares methods	Shank	0.02816166	0.05453870
	Prony	0.02783890	0.05366054
	Pade	0.08187592	0.15415749
Suboptimal methods	Kung	0.16565446	0.33124686
	Kimura	0.01845821	0.03450080
	MS	0.03792555	0.02040313
Optimal Hankel methods	OPH	0.16367153	0.32733688
	OPHd	0.01631812	0.02778568
	OPHD	0.01635045	0.02900526
	CF	0.02067489	0.04092900

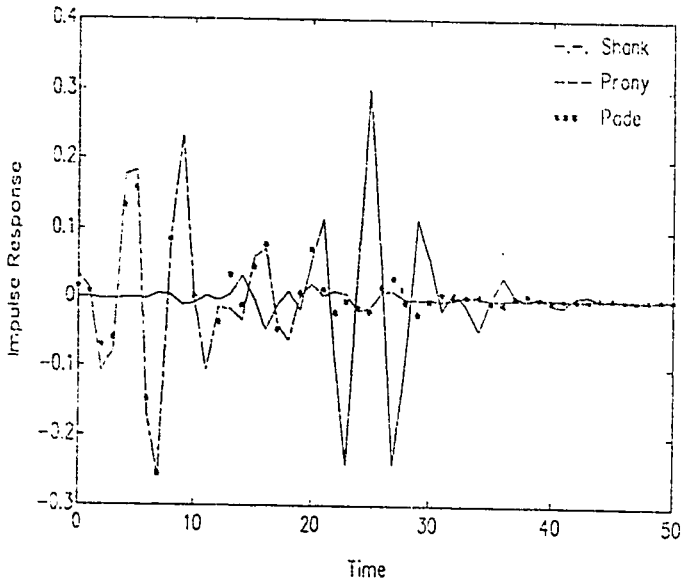


Fig. 5.44a: Impulse response of least squares methods using two-sided approximation technique with r = 8.

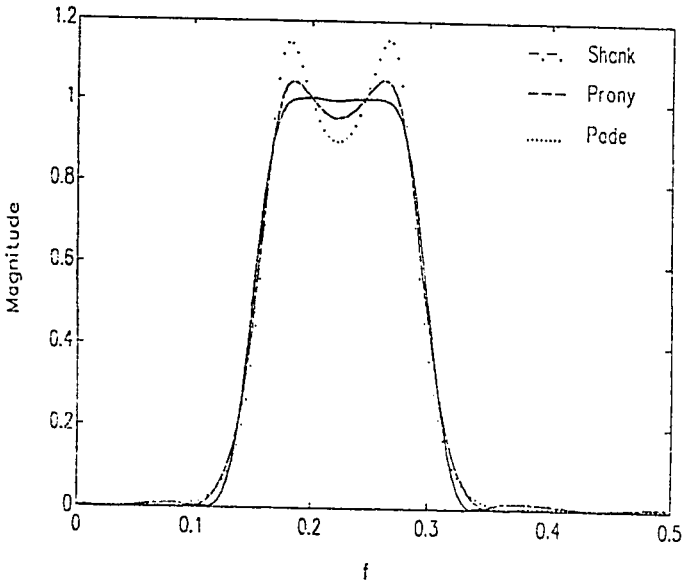


Fig. 5.44b: Magnitude response of least squares methods using two-sided approximation technique with r = 8.

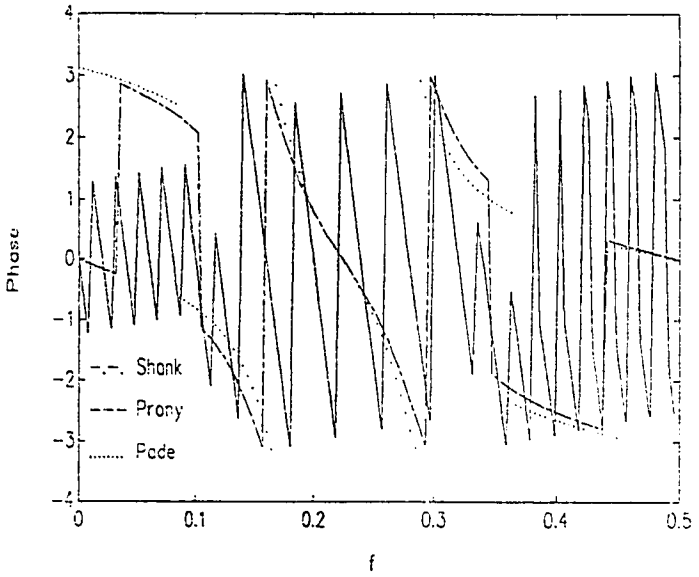


Fig. 5.44c: Phase response of least squares methods using two-sided approximation technique with r = 8.

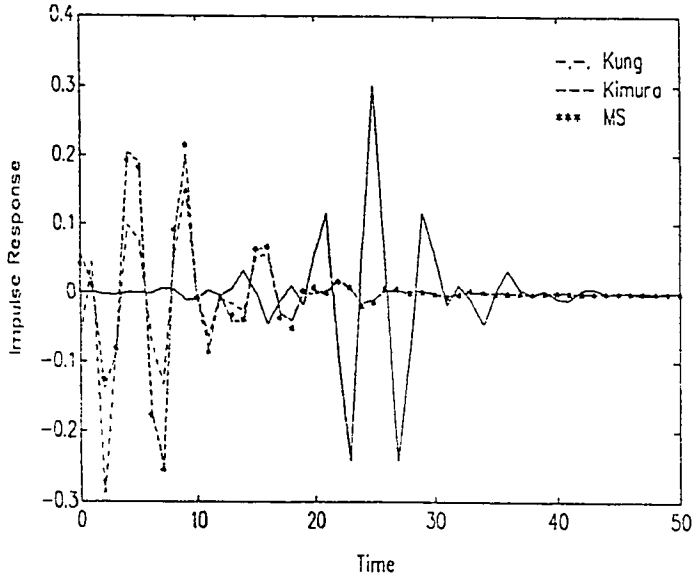


Fig. 5.45a: Impulse response of suboptimal methods using two-sided approximation technique with r = 8.

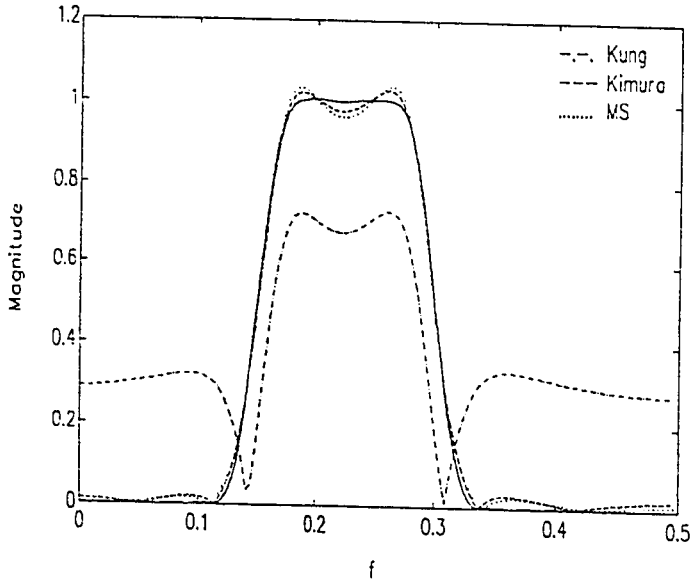


Fig. 5.45b: Magnitude response of suboptimal methods using two-sided approximation technique with r = 8.

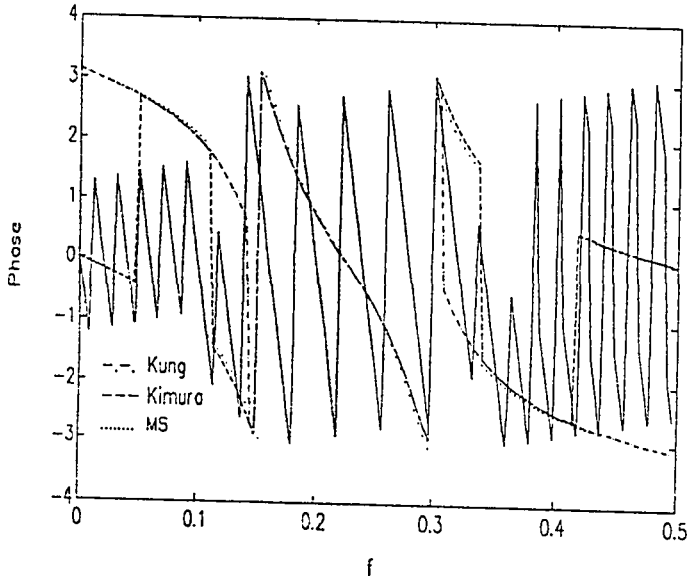


Fig. 5.45c: Phase response of suboptimal methods using two-sided approximation technique with r = 8.

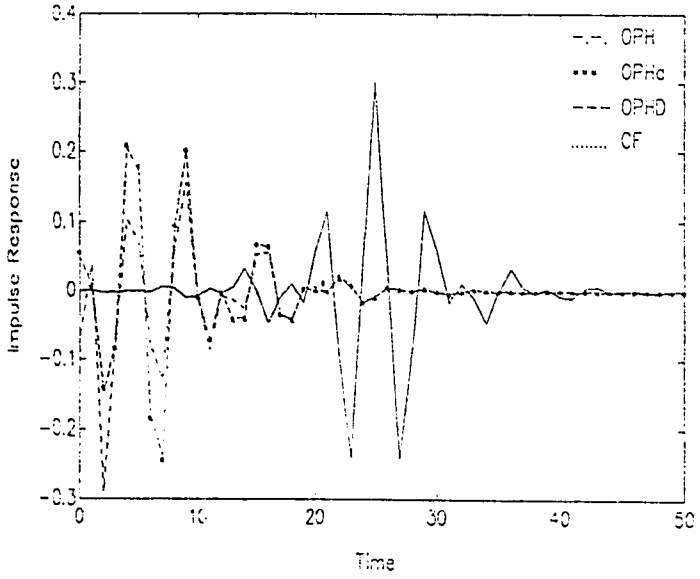


Fig. 5.46a: Impulse response of optimal Hankel methods using two-sided approximation technique with r = 8.

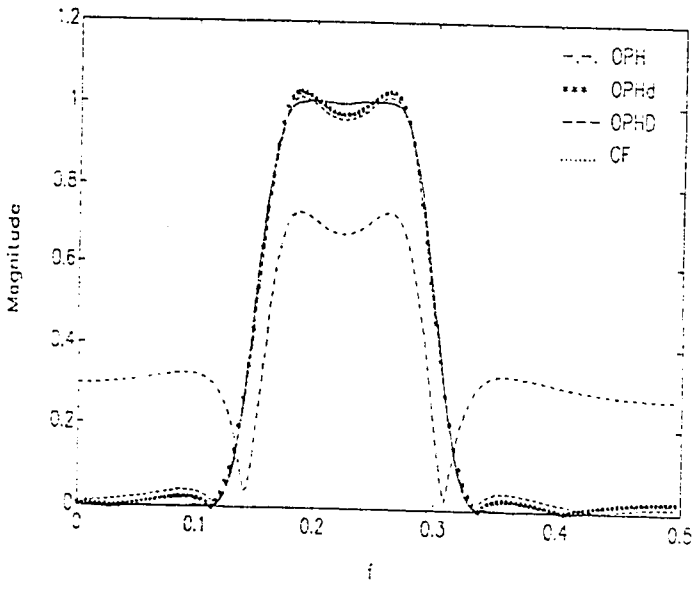


Fig. 5.46b: Magnitude response of optimal Hankel methods using two-sided approximation technique with r = 8.

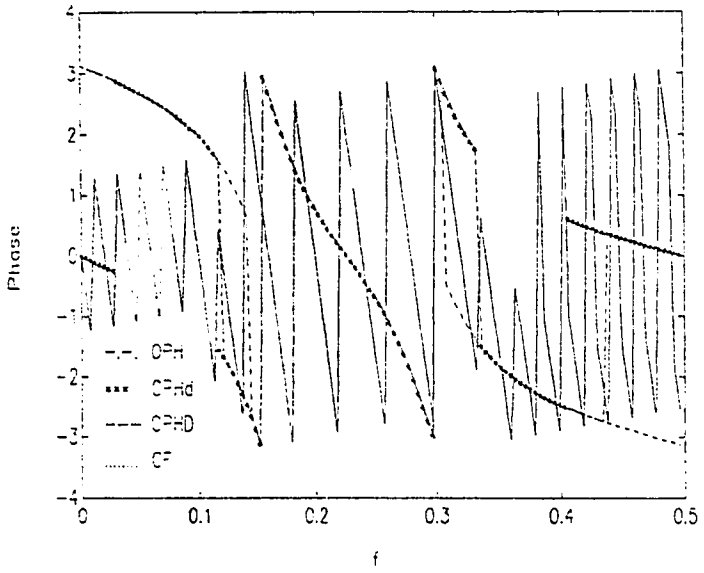


Fig. 5.46c: Phase response of optimal Hankel methods using two-sided approximation technique with r = 8.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The main conclusions of this thesis is listed below.

1) Optimal Hankel methods and suboptimal Hankel methods are considered to be superior compared to least squares methods from digital filter design point of view.

2) Suboptimal Hankel methods, namely: Kung method, Kimura method, and MS method sometimes give IIR digital filter designs which are comparable to the designs obtained by optimal Hankel methods.

3) The Hankel methods which omit the D-term: Kung method and OPH method suffer from a serious error when h_0, the DC term of the impulse response, is comparatively large. When h_0 is very small, the error is negligible.

4) The performance of OPH method is greatly improved when the D-term is forced to it (OPHd method) provided that the D-term is large. However, when the D-term is small, OPH and OPHd methods give almost the same design.

5) When a designer is only concerned about the pass-band of a desired LPF, MS method is the best method that could be applied.

6) The frequency response and impulse response of the IIR digital filter designed using CF method are mostly identical to the response of the IIR digital

filter designed using OPHD method which indicates the optimality of this method. However, it has a main disadvantage where it could fail in finding a suitable parametric form to the impulse response using Prony method.

7) OPHD method, OPHd method and Kimura method are the best Hankel methods for IIR digital filter design and approximation in general.

8) Hankel methods and balanced approximation methods give a very efficient approximation to the phase response in the pass-band. However, this is not the case for the stop-band where the approximation is generally poor.

9) Although OPHD method has a lower error bound in frequency domain than OPH and OPHd methods, OPH and OPHd methods could achieve a lower error which indicates the optimality of these methods.

10) Two-sided approximation technique is a very powerful tool available for a designer to design or approximate FIR or IIR digital filters (with symmetric impulse response) when the magnitude response is only concerned.

11) Those methods which suffer from the D-term are not efficient when two-sided approximation technique is applied to them.

6.2 Recommendations

1) Apply frequency weighted Hankel methods to digital filter design and compare the results to other methods.

2) Expand two-sided approximation technique to include non-symmetric impulse response.

APPENDIX A

A.1 Pade Approximation

Pade approximation technique for IIR digital filter design. The input arguments are the impulse response h, the numerator order m, and the denominator order n. The output is given as the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf] = pade(h, m ,n)
format long
hh = h; l = length(hh); h = h(1:m+n+1);
K = length(h) - 1;
M = m; N = n;
H = tril(toeplitz(h));
% K+1 by N+1
if (K > N)
        H(:,(N+2):(K+1)) = [ ];
end
% Partition H matrix
H1 = H(1:(M+1),:);      % M+1 by N+1
h1 = H((M+2):(K+1),1);  % K-M by 1
H2 = H((M+2):(K+1),2:(N+1));    % K-M by N

% Calculation of numerator coefficients b and denominator coefficients a.
a = [1; -H2\h1].';
b = a*H1.';
HH=H2'*H2;

% Calculation of LSE
x = [1 zeros(1:l-1)]; y = filter(b,a,x);
for i=1:l ; err(i) = hh(i)-y(i); end;
lse = norm(err,2);

% Calculation of L∞  error norm
[magl,w]=freqz(b,a,256,'whole'); mag = abs(magl); ph = angle(magl);
magr = fft(hh,256);
for i = 1:256; errf(i) = magr(i)-magl(i); end
```

```
linf = norm(errf,inf);
end
```

A.2 Prony Method

Least squares method (Prony method) for IIR digital filter design. The input arguments are the impulse response h, the numerator order m, and the denominator order n. The output is given as the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf] = prony(h, m ,n)

K = length(h) - 1;
M = m; N = n;
H = tril(toeplitz(h));
% K+1 by N+1
if (K > N)
        H(:,(N+2):(K+1)) = [ ];
end
% Partition H matrix
H1 = H(1:(M+1),:);   % M+1 by N+1
h1 = H((M+2):(K+1),1);       % K-M by 1
H2 = H((M+2):(K+1),2:(N+1));       % K-M by N

% Calculation of numerator coefficients b and denominator coefficients a
a = [1; -(H2'*H2)\(H2'*h1)].';
b = a*H1.';

% Calculation of LSE
x=[1 zeros(1:K)]; y=filter(b,a,x);
for i=1:K+1; err(i)=h(i)-y(i); end
lse=norm(err,2);

% Calculation of L∞ error norm.
[magl,w]=freqz(b,a,256,'whole'); mag=abs(magl); ph=angle(magl);
magr=fft(h,256);
for i=1:256; errf(i)=magr(i)-magl(i); end
linf=norm(errf,inf);

end
```

A.3 Shank method

Least squares method (Shank) for IIR filter design. The input arguments are the impulse response h, the numerator order m, and the denominator order n. The output is given as the numerator coefficients b, denominator coefficients a, LSE, and L_∞.

Program:

```
function [b,a,lse,linf]=lsqr(h,m,n)

% The following steps calculate a, the denominator coefficients.

l=length(h); x=h(n+2:l)' ;
for i=1:l-n-1
for j=1:m; y(i,j)=h(n-j+i+1); end
end
a=inv(y'*y)*y'*x;

% The following steps calculate b, the numerator coefficients.

g1(1)=1;
for i=2:l
sum=0;
for j=1:m
if (i-j)>0; s=a(j)*g1(i-j);
else; break
end
sum=sum+s; end
g1(i)=sum; end

for i=1:l
for j=1:n+1
 if (i-j) < 0;  g(i,j)=0;
 else;  g(i,j)=g1(i-j+1);
 end ;
end
end

b=inv(g'*g)*g'*h'; a=[1;-a];

% The following steps calculate  L∞ error norm and LSE.
```

```
x=[1 zeros(1:l-1)]; y=filter(b,a,x);
for i=1:l ; err(i)=h(i)-y(i); end
lse=norm(err,2);

[magl,w]=freqz(b,a,256,'whole'); mag=abs(magl); ph=angle(magl);
magr=fft(h,256);
for i=1:256; errf(i)=magr(i)-magl(i); end
linf=norm(errf,inf);

end
```

A.4 Kung Method

Kung's method is applied to IIR digital filter design where the impulse response h is required only as an input. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm

Program:

```
function [b,a,lse,linf]=kung(h);
format long
h0=h(1); h=h(2:length(h)); k=length(h); l=256;
han=hankel(h); [u,s,v]=svd(han);

% The following two steps enable the designer to choose the order of the
% reduced model.

sv=diag(s)
n=input(' n ? ')

% The following steps calculate the submatrices required to find the reduced
% model.

sig1=s(1:n,1:n); sig2=s(n+1:k,n+1:k);
u1=u(:,1:n); u2=u(:,n+1:k);
v1=v(:,1:n)'; v2=v(:,n+1:k)';
u1s1=u1*sqrtm(sig1); u2s2=u2*sqrtm(sig2);

% The following steps calculate the one-row shift upward matrix.
```

```
shf1=u1s1(2:k,:); shf2=u1s1(1,:); shf3=[shf1 ; zeros(1,n)];

% The following steps calculate the approximate reduced realization
% (a11,b1,c1).

a11=inv(sqrtm(sig1))*u1'*shf3; db1=sqrtm(sig1)*v1;
b1=db1(:,1);
c1=u1s1(1,:);

% The reduced model as a rational function H(z)=b(z)/a(z).

[b,a]=ss2tf(a11,b1,c1,0,1);
[magl,w]=freqz(b,a,l,'whole');
mag=abs(magl); ph=angle(magl);

% Calculation of L∞ error norm and LSE.

h=[h0 h];
magr=fft(h,l);
for i=1:l ; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf);

x=[1 zeros(1:k)]; y=filter(b,a,x);
for i=1:k+1; err(i)=h(i)-y(i); end
lse=norm(err);

end
```

A.5 Kimura and Honoki Method

Kimura and Honoki method for IIR filter design where the impulse response h is required only as an input. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf]=balap1(h);
format long
h0=h(1);h=h(2:length(h)); k=length(h); l=256;
```

```
han=hankel(h); [u,s,v]=svd(han);

% The following two steps enable the designer to choose the order of the
% reduced model.

sv=diag(s)
n=input(' n ? ')

% The following steps calculate the submatrices required to find the reduced
% model.

sig1=s(1:n,1:n); sig2=s(n+1:k,n+1:k); u1=u(:,1:n); u2=u(:,n+1:k);
v1=v(:,1:n)'; v2=v(:,n+1:k)';
u1s1=u1*sqrtm(sig1); u2s2=u2*sqrtm(sig2);

% The following steps calculate the one-row shift upward matrix.

shf1=u1s1(2:k,:); shf2=u1s1(1,:); shf3=[shf1 ; zeros(1,n)];

% The following steps calculate the approximate reduced realization
% (a11,b1,c1).

a11=inv(sqrtm(sig1))*u1'*shf3; db1=sqrtm(sig1)*v1;
b1=db1(:,1);
c1=u1s1(1,:);

% The reduced model as a rational function H(z)=b(z)/a(z).

[b,a]=ss2tf(a11,b1,c1,h0,1);
[magl,w]=freqz(b,a,l,'whole'); mag=abs(magl); ph=angle(magl);

% Calculation of L∞ error norm and LSE.

h=[h0 h]; magr=fft(h,l);
for i=1:l; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf);

x=[1 zeros(1:k)]; y=filter(b,a,x);
for i=1:k+1; err(i)=h(i)-y(i); end
lse=norm(err);

end
```

A.6 MS Method

Minimum sensitivity method for IIR filter design where the impulse response h is required only as an input. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf]=balap2(h)
format long
l=256; h0=h(1); h=h(2:length(h)); m=length(h);
han=hankel(h); [u,s,v]=svd(han);

% The following two steps enable the designer to choose the order of the
% reduced model.

sv=diag(s)
r=input(' r ? ')

% The following steps calculate the submatrices required to calculate the
% balanced model and the reduced model of the original system.

sqr=sqrtm(s); us=u*sqr; us1=us(2:m,:); us2=[us1 ; zeros(1,m)];
a=inv(us)*us2;
bt=sqr*v';
b=bt(:,1);
c=us(1,:);
a11=a(1:r,1:r); a12=a(1:r,r+1:m); a21=a(r+1:m,1:r); a22=a(r+1:m,r+1:m);
b1=b(1:r,:); b2=b(r+1:m,:);
c1=c(:,1:r); c2=c(:,r+1:m);

% The following steps calculate the reduced balanced model.

t=inv((eye(a22))-a22);
ar=a11+(a12*t*a21);
br=b1+(a12*t*b2);
cr=c1+(c2*t*a21);
dr=h0+c2*t*b2;

% The following steps calculate the reduce balanced model which has the
% minimum sensitivity property to both parameter variation and roundoff noise.
% balap22 is a subroutine to calculate P, the nonsingular transformation matrix.
```

```
g1=diag(s); del=sum(g1(1:r));
g2=sqrt(del/r);

rt=balap22(ar,br,cr,r);
t=g2*rt';
art=inv(t)*ar*t;
brt=inv(t)*br;
crt=cr*t;

% The following steps calculate the L∞ error norm and LSE.

[b,a]=ss2tf(art,brt,crt,dr,1); [magl,w]=freqz(b,a,l,'whole');
mag=abs(magl); ph=angle(magl);
h=[h0 h];magr=fft(h,l);

for i=1:l; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf)

x=[1 zeros(1:m)]; y=filter(b,a,x);
for i=1:m+1; err(i)=h(i)-y(i); end
lse=norm(err);

end
%-------------------------------------------------------------------------------
%-------------------------------------------------------------------------------

%   Subroutine balap22 which calculates the nonsingular transformation matrix
%    P for minimum sensitivity method.

function rt=balap22(ar,br,cr,r)
format long
n=r; sum1=0 ; sum2=0;
for i=0:70
k=ar^i*br*br'*(ar^i)'; sum1=sum1+k; w=(ar^i)'*cr'*cr*ar^i;
sum2=sum2+w;
end

q=sum1*sum2; m=eig(q); u=sqrt(m(1:n)); u=n*u(1:n)/sum(u);
z=diag(u);

z1=z; z2=z; rt=eye(z);
for i=n:-1:2
```

```
for i=n:-1:2
r1=rt;
r=eye(z);
for j=i-1:-1:1
 if z(j,j)>1
 u1=(z(j,j)-1)/(z(j,j)-z(i,i)); u2=(1-z(i,i))/(z(j,j)-z(i,i));
 r(j,j)=sqrt(u1); r(i,i)=r(j,j); r(i,j)=sqrt(u2); r(j,i)=-r(i,j);
 break
 end
end
rt=r*rt; z=rt*z2*rt';
end
z1=rt*z1*rt';

end
```

A.7 OPH Method

Optimal Hankel method (OPH) without a D-term applied to IIR filter design where the impulse response h is required only as an input. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf]=OPH(h)
format long
l=256; h0=h(1);h=h(2:length(h)); n=length(h);
c=hankel(h); [u,s,v]=svd(c);

% The following steps enable the designer to choose the order of the reduced
% model.

sv=diag(s)
r=input(' r ? ');

% The following steps calculate the optimal Hankel reduced model.

m=u(:,r+1);
for i=1:n-1
sm=0;
```

```
for j=1:i
k=h(i-j+1)*m(n-j+1); sm=sm+k;
end
cof(i)=sm;
end
cof=[0 cof]; m=m(n:-1:1);
[r,p,k]=residue(cof,m);

% The following steps calculate the causal part of the optimal Hankel reduced
% model in the form of a rational function b(z)/a(z).

j=1;
for i=1:n-1
if abs(p(i))<1 ; p1(j)=p(i); r1(j)=r(i); j=j+1; end
end

[b,a]=residue(r1,p1,k); b=[0 b];

% The following steps calculate  L∞ error norm and LSE.

h=[h0 h]; x=[1 zeros(1:n)]; y=filter(b,a,x);
for i=1:n+1; err(i)=h(i)-y(i); end
lse=norm(err);

[magl,w]=freqz(b,a,l,'whole'); mag=abs(magl); ph=angle(magl);
magr=fft(h,l);
for i=1:l; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf);

end
```

A.8 OPHd Method

Optimal Hankel method (OPH) with the D-term forced to it applied to IIR digital filter design where the impulse response h is required only as an input. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf]=OPH_d(h)
```

```
format long
l=256; h0=h(1);h=h(2:length(h)); n=length(h);
c=hankel(h); [u,s,v]=svd(c);

% The following steps enable the designer to choose the order of the reduced
% model.
sv=diag(s)
r=input(' r ? ');

% The following steps calculate the optimal Hankel reduced model.

m=u(:,r+1);
for i=1:n-1
sm=0;
for j=1:i
k=h(i-j+1)*m(n-j+1); sm=sm+k;
end
cof(i)=sm;
end

cof=[0 cof]; m=m(n:-1:1);
[r,p,k]=residue(cof,m);

% The following steps calculate the causal part of the optimal Hankel reduced
% model in the form of a rational function b(z)/a(z).

j=1;
for i=1:n-1
if abs(p(i))<1 ; p1(j)=p(i); r1(j)=r(i); j=j+1; end
end

[b,a]=residue(r1,p1,k); b=[0 b];
[ad,bd,cd,dd]=tf2ss(b,a); dd=h0;
[b,a]=ss2tf(ad,bd,cd,dd,1);

% The following steps calculate  L∞ error norm and LSE.

h=[h0 h]; x=[1 zeros(1:n)]; y=filter(b,a,x);
for i=1:n+1; err(i)=h(i)-y(i); end
lse=norm(err);

[magl,w]=freqz(b,a,l,'whole'); mag=abs(magl); ph=angle(magl);
```

```
magr=fft(h,l);
for i=1:l; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf);

end
```

A.9 Glover Method

This method gives an optimal Hankel design to IIR digital filters. It requires the impulse response h as an input only. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf]=glover(h);
n=length(h); num=h ; dem=[1 zeros(1:n-1)];
[a,b,c,d1]=tf2ss(num,dem);

% The following steps convert the given system from discrete to continuous.

invm=inv(eye(a)+a);
ac=invm*(a-eye(a));
bc=sqrt(2)*invm*b;
cc=sqrt(2)*c*invm;
dc=d1-(c*invm*b);

% The following steps find optimal Hankel approximation using Glover's
% algorithm.

sys=pck(ac,bc,cc,dc); [sysb,sv]=sysbal(sys,0.0);
sv
k=input(' k ? ');
[sysred,sysanti,siganti]=hankmr(sysb,sv,k,'d');
 [ac,bc,cc,dc]=unpck(sysred);

% The following steps find the discrete version of the optimal Hankel
% approximation.
```

```
if (k = = 1); idn=1; else ; idn=eye(ac); end
invm=inv(idn-ac);
ad=(ac+idn)*invm;
bd=sqrt(2)*invm*bc;
cd=sqrt(2)*cc*invm;
dd=dc+cc*invm*bc;

% The following steps calculate L∞ error norm and LSE.

[b,a]=ss2tf(ad,bd,cd,dd,1); [magl,w]=freqz(b,a,256,'whole');
mag=abs(magl); ph=angle(magl);
x=[1 zeros(1:n-1)]; y=filter(b,a,x);
for i=1:n; dif(i)=h(i)-y(i); end
lse=norm(dif);

magr=fft(h,256);
for i=1:256 ; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf);

end
```

A.10 CF Method

CF method for IIR digital filter deign. It requires the impulse response h, numerator order m, and denominator order n as an input. It returns back the numerator coefficients b, denominator coefficients a, LSE, and L_∞ error norm.

Program:

```
function [b,a,lse,linf]=cf(h,m,n);
k=length(h); c=m-n+1; l=256;

% The following steps calculate the required Hankel matrix.

for i=1:k-c
for j=1:k-c
 t=i+j+c-1;
 if(t<=k & t>0); han(i,j)=h(t); else; han(i,j)=0; end
end
```

```
end

% The following steps calculate the optimal chebyshev approximation.

[u,s,v]=svd(han); sig=diag(s);
fk=fft(h,l); un=u(:,n+1); vn=v(:,n+1); fun=fft(un,l); fvn=fft(vn,l); fvb=conj(fvn);
j=sqrt(-1);
for i=0:l-1
w(i+1)=2*pi*i/l; z=exp(j*w(i+1));
r(i+1)=fk(i+1)-sig(n+1)*(z.^(-c))*(fun(i+1)/fvb(i+1));
end

% The following steps calculate the impulse response of the optimal
% approximation in Chebyshev sense and then a near optimal rational
% approximation is found using Prony method after getting red of the
% noncausal part of the impulse response.

rt=ifft(r);
if(c >= 0); rcf=[rt(1:l/2) zeros((l/2)+1:l)]; [b,a]=prony(rcf,m,n);
else;
x=abs(c);
rcf=[zeros(1:x) rt(1:(l/2)-x) zeros((l/2)-x+1:l-x)] ; [b1,a]=prony(rcf,n-1,n);
for i=1:m+1
b(i)=b1(i-c);
end
end

% The following steps calculate L∞ error norm and LSE.

x=[1 zeros(1:k-1)]; y=filter(b,a,x);
for i=1:k
err(i)=h(i)-y(i);
end
lse=norm(err);

[magl,w]=freqz(nr,dr,l,'whole'); mag=abs(magl); ph=angle(magl);
magr=fft(h,l);
for i=1:l ; dif(i)=magr(i)-magl(i); end
linf=norm(dif,inf);

end
```

A.11 Two-Sided Approximation Technique

This program enables the designer to apply two-sided approximation technique to any time domain IIR filter design method. It requires the impulse response h as input. The impulse response h should be symmetric or antisymmetric. The program returns back the filter coefficients as c(z)/d(z), LSE, and L_∞ error norm for *magnitude response*.

Program:

```
function [c,d,lse,linf]=twosd(h)
format long
l=length(h); ll=(l+1)/2; h1=[h(ll)/2 h(ll+1:l)];

disp (' ')
disp (' The following step enables the designer to choose the filter order')
disp (' (m,n) where m is the numerator order and n is the denominator order.')
disp (' The final order of the designed filter will be (2m,2n).')
disp (' ')

m=input(' m ? '); n=input(' n ? ');

disp (' ')
disp (' For optimal and suboptimal Hankel methods m=n=r where r is the')
disp (' order of the reduced model. The program will force m and n to ')
disp (' be equal to r. Please press any key to continue.')
disp (' ')
pause

disp (' ')
disp (' For the value of s: put 0 if h is symmetric')
disp ('                     put 1 if h is antisymmetric')
disp (' ')

s=input(' s ? ');

% In the following step any IIR filter design method could be used instead
% of Pade method provided that input arguments are adjusted. This step
% provides us with a rational approximation to the causal part of the impulse
% response.

[b11,a11]=pade(h1,m,n);
```

```
m=length(b11)-1; n=length(a11)-1; k=abs(m-n);
b22=b11(m+1:-1:1); a22=a11(n+1:-1:1);

for i=1:m+n+1
bb(i)=0; aa(i)=0;
for j=1:i
if (j <= m+1 & i-j <= n); bb(i)=bb(i)+b11(j)*a11(n-i+j+1); end
if (j <= n+1 & i-j <= m); aa(i)=aa(i)+a11(j)*b11(m-i+j+1); end
end
end

% The following steps calculate c(z), the numerator coefficients of
% the designed filter, when h is symmetric.

if ( s == 0)
if (m <= n)
for i=1:m+n+k+1
if (i<k+1); c(i)=bb(i);
elseif (i>m+n+1); c(i)=aa(i-k);
else; c(i)=bb(i)+aa(i-k); end
end

else
for i=1:m+n+k+1
if (i<k+1); c(i)=aa(i);
elseif (i>m+n+1); c(i)=bb(i-k);
else; c(i)=aa(i)+bb(i-k); end
end
end

% The following steps calculate c(z), the numerator coefficients of the
% designed filter, when h is antisymmetric.

else % s=1
if (m <= n)
for i=1:m+n+k+1
if (i<k+1); c(i)=bb(i);
elseif (i>m+n+1); c(i)=-aa(i-k);
else; c(i)=bb(i)-aa(i-k); end
end

else
```

```
for i=1:m+n+k+1
if (i<k+1); c(i)=-aa(i);
elseif (i>m+n+1); c(i)=bb(i-k);
else; c(i)=bb(i-k)-aa(i); end
end
end
end

% The following step calculate d(z), the denominator coefficients of the
% the designed filter.
d=conv(a11,a11);

% The following steps calculate LSE and L∞ error norm for the
% magnitude response of the designed filter.

x=[ 1 zeros(1:l-1)]; y=filter(c,d,x);
for i=1:l; err(i)=h(i)-y(i); end;
lse=norm(err);

[mt,w]=freqz(c,d,128); magt=abs(mt); m=fft(h,256); mag=abs(m);
for i=1:128; err(i)=mag(i)-magt(i); end
linf=norm(err,inf);

end
```

NOMENCLATURE

LSE	least-squares error
h_0	the DC-term of the impulse response, h(0)
BPF	band-pass filter
HPF	high-pass filter
LPF	low-pass filter
LPH	linear phase
NLPH	non-linear phase
LPHC	linear phase characteristic
FIR	finite impulse response
IIR	infinite impulse response
O,G	observability and controllability respectively
W_o, W_c	observability grammian and controllability grammian respectively
H	Hankel matrix unless otherwise stated
H(f(z))	Hankel matrix of the transfer function f(z)
$\|A\|_s$	spectral norm of A: $\|A\|_s = \sigma_{max}[A^T A]$
$\|f(z)\|_H$	Hankel orm of f(z): $\|f(z)\|_H = \sigma_{max}[H(f(z))^T H(f(z))] = \sigma_{max}[W_c W_o]$
$\|f(z)\|_\infty$	Chebyshev norm of f(z): $\|f(z)\|_\infty = \underset{0\le\omega\le 2\pi}{\text{ess sup}} \|f(e^{j\omega})\|$

REFERENCES

1. L. R. Rabiner and B. Gold, Theory and Application of Digital Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1975.

2. A. G. Evans and R. Fischl, "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters," IEEE Trans. on Audio and Electroacoustics, AU-21, No. 1, 61-65, Feb., 1973.

3. R. Fischl, "Optimal L_p- approximation of Prescribed Impulse Response Functions on a Finite Point Set," in Proc. IEEE Int. Symp. Circuit Theory, 1970, pp. 155-156.

4. C. S. Burrus, T. W. Parks, and T. B. Watt, Jr., "A Digital Parameter-Identification Technique Applied to Biological Signals," IEEE Trans. Bio-Med. Eng., vol. BME-18, pp. 35-37, Jan. 1971.

5. F. Brophy and A. C. Salazar, "Recursive Digital Filter Synthesis in the Time Domain," IEEE Trans. Acoust., Speech, and Signal processing, vol. ASSP-22, No. 1, 45-55, Feb., 1974.

6. C. K. Chui and A. K. Chan, "A Two-Sided Rational Approximation Method for Recursive Digital Filtering," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-27, pp. 141-145, 1979.

7. G. Forbenius, "Über Relationen Zwischen den Naherungsbruche von Potenzreichen," J. fur reine and angew. Math., vol. 90, pp. 1-17, 1881.

8. F. Brophy and A. C. Salazar, "Considerations of Pade Approximant Technique in the Synthesis of Recursive Digital Filters," IEEE Trans. Audio Electroacoust., vol. AU-21, pp. 500-505, Dec. 1973.

9. C. S. Burrus and T. W. Parks, "Time Domain Design of Recursive Digital Filters," IEEE Audio Electroacoust., vol. AU-18, pp. 137-141, June 1970.

10. R. Hastings-James and S. Mehra, "Extensions of the Pade-Approximant Technique for the Design of Recursive Digital Filters," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-25, pp. 501-509, Dec. 1977.

11. C. K. Chui and A. K. Chan, "A New Approach to Causal Filter by Pade Approxi mants," ICASSP, Denver, vol. 1, pp. 264-267, 1980.

12. L. E. McBride, Jr., H. W. Schaefgen, and K. Stegilitz, "Time-Domain Approximaion by Iterative Methods," IEEE Trans. Circuit Theory, vol. CT-13, pp. 381-387, Dec. 1966.

13. J. L. Shanks, "Recursion Filters for Digital Processing, "Geophysics, vol. XXXII, no. 1, pp. 33-51, Feb. 1967.

14. J. G. Proakis and D. G. Manolakis, Digital Signal Processing Principles, Algorithms, and Applications. 2nd ed., Macmillan Publishing Co., New york, 1992.

15. M. H. Gutknecht, J. O. Smith, and L. N. Trefethen, "The Caratheodory-Fejer Method for Recursive Digital Filter Design," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-31, pp. 1417-1426, Dec. 1983.

16. T. Takagi, "On an Algebraic Problem Related to an Analytic Theorem of Caratheodory and Fejer and on an Allied Theorem of Landau" and "Remarks on an Algebraic Problem," Japan J. Math., vol. 1, pp. 83-93, 1924, and vol. 2, pp. 13-17, 1925.

17. C. Caratheodory and L. Fejer, " Über den Zusammenhang der Extremen von Harmonischen Funktionen mit ihrer Koeffizienten und über den Picard-Landauschen Satz.," Rend. Circ. Mat. Palermo, vol. 32, pp. 218-239, 1911.

18. M. Gutknecht and L. N. Trefethen, "Recursive Digital Filter Design by the Caratheodory-Fejer Method," Dep. Comput. Sci., Stanford Univ., Stanford, CA, Numer. Anal. ms. NA-80-01, 1980.

19. L. N. Trefethen, "Rational Chebyshev Approximation on the Unit Disk," Numer. Math., vol. 37, pp. 297-320, 1981.

20. M. Gutknecht, "Rational Caratheodory-Fejer Approximation on a Disk, a Circle, and an Interval," J. Approx. Theory, vol. 41, pp. 257-278, 1984.

21. L. N. Trefethen and M. H. Gutknecht, "The Caratheodory-Fejer Method for Real Rational Approximation," SIAM J. Numer. Anal., vol. 20, pp. 420-436, Apr. 1983.

22. B. C. Moore,"Singular Value Analysis of Linear Systems," Proc.1978 IEEE CDC, pp. 66-73, 1978.

23. C. T. Mullis and R. A. Roberts, "Synthesis of Minimum Roundoff Noise Fixed Point Digital Filters," IEEE Trans. Circ. Sys., vol. CAS-23, pp. 551-562, Sept. 1976.

24. L. Pernebo and L. M. Silverman, "Model Reduction via Balanced State-Space Representations," IEEE Trans. Automat. Contr., vol. AC-27, pp. 382-387,1982.

25. A. M. Davidson, "Balanced Systems and Model Reduction," Electron. Lett., vol. 22, pp. 531-532, 1986.

26. P. T. Kabamba, "Balanced Gains and their Significance for L^2 Model Reduction," IEEE Trans., vol. AC-30, pp. 690-693,1985.

27. S. Kung, "A New Identification and Model Reduction Algorithm via Singular Value Decompositions," In Proc. 12th Annu. Asilomar Conf. Circuits, Syst., Comput., pp. 705-714, Nov. 1978.

28. M. Reuter, "Hankel Approximation Methods for the Design of IIR Filters via Singular Value Decomposition," M.S Thesis, University of Illinois, 1986.

29. G. M. Pitstick, J. R. Cruz, and R. J. Mulholland, "Approximate Realization Algorithms for Truncated Impulse Response Data," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1583-1587, Dec. 1986.

30. J. M. Mendel, "Minimum-Variance and Maximum-Likelihood Recursive Wave-Shaping," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-31, pp. 599-604, June 1983.

31. L. M. Silverman and M. Bettayeb, "Optimal Approximation of Linear Systems," in Proc. JACC, San Francisco, CA, 1980.

32. M. Bettayeb, "Approximation of Linear Systems: New Approaches Based on Singular Value Decomposition," Ph.D. Dissertation, University of Southern California, Los Angeles, Aug. 1981.

33. K. Glover, "All Optimal Hankel-Norm Approximation of Linear Multivariable Systems and their L^∞–Error Bounds," Int. J. Control, vol. 39, No. 6, pp. 1115-1193, 1984.

34. K. V. Fernando and H. Nicholson, "Singular Perturbational Approximations for Discrete-Time Balanced Systems," IEEE Tarns. Automat. Contr., AC-28, pp. 240-242, Feb. 1983.

35. W. J. Lutz and S. L. Hakimi, "Design of Multi-Input Multi-Ouput Systems with Minimum Sensitivity," IEEE Trans. Circ. Sys., vol. CAS-35, pp. 1114-1122, 1988.

36. L. Thiele, "Design of Sensitivity and Round-off Noise Optimal State-Space Discrete Systems," Int. J. Circuit Theory Appl., vol. 12, pp. 39-46, 1984.

37. __, "On the Sensitivity of Linear State-Space Systems,"IEEE Trans. Circ. Sys., vol. CAS-33, pp. 502-510, 1988.

38. U. M. Al-Saggaf, "On Model Reduction and Control of Discrete-Time Systems," Ph.D dissertation, Information Systems Laboratory, Dept. of Electrical Engineering, Stanford University, June 1986.

39. K. Heckelmann and R. Unbehauen, "Approximation of the Frequency Response of a FIR Digital Filter by an IIR Filter," in Proc. European Conf. Circuits Theory and Design, ECCTD 87, Paris, 1987, pp. 477-482.

40. J. B. Bednar and W. A. Coberly, "Order Selection for and Design of IIR Filters," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-30, pp. 211-216, Apr. 1982.

41. J. B. Bednar, "On the Approximation of FIR by IIR Digital Filters," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-31, pp. 28-34, Feb. 1983.

42. H. Kimura and Y. Honoki, "Balanced Approximation of Digital FIR Filter with Linear Phase Characteristic," in Proc. ISCAS 85, Japan, pp. 283-286.

43. B. Beliczynski, I. Kale, and G. D. Cain, "Approximation of FIR by IIR Digital Filters: An Algorithm Based on Balanced Model Reduction," IEEE Trans. on Sig. Processing, vol. 40, No. 3, pp. 532-542, Mar. 1992.

44. V. Sreeram and P. Agathoklis, "Design of Linear-Phase IIR Filters via Impulse-Response Gramians," IEEE Trans. Signal Processing, vol. 40, No. 2, Feb. 1992.

45. C. Chen, "Pade Approximants of Noncausal Digital Filters,"Journal of the Franklin Institute, vol. 310, No. 4/5, pp. 209-213, Oct./Nov. 1980.

46. J. J. Kormylo and V. K. Jain, "Two Pass Recursive Digital Filter with Zero Phase Shift," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-22, pp. 384-387, Oct. 1974.

47. R. Czarnach, "Recursive Processing by Noncausal Digital Filters," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, pp. 363-370, June 1982.

48. C. K. Chan and C. F. Chen, "Sample-by-Sample Approach to Recursive Noncausal Filtering," Electro. Lett., vol. 20, No. 4, pp. 172-174, Feb. 1984.

49. G. Cortelazo and M. R. Lightner, "The Use of Multiple Criterion Optimization for Frequency Domain Design of Noncausal IIR Filters," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 126-135, Feb. 1985.

50. C-K. Chan, "Noncausal Digital Filters with Antisymmetrical Impulse Response," ICASSP, Tokyo, pp. 2611-2614, 1986.

51. A. J. Laub, M. T. Heath, C. C. Paige and R. C. Ward, "Computation of System Balancing Transformations and other Applications of Simultaneous Diagonalization Algorithms," IEEE Trans. on Autom. Cont., vol. AC-32, No. 2, pp. 115-121, 1987.

52. B. L. Ho and R. E. Kalman, "Effective Construction of Linear State-Variable Models from Input/Output Data," in Proc. 3rd Ann. Allerton Conf., Monticello, I11., pp. 449-459, Oct. 1965

53. B. S. Chen, B. W. Chiou, and S. C. Peng, "Minimum Sensitivity IIR Filter Design Using Principal Component Approach," IEE Proc. G, vol.138, pp. 474-482, Aug. 1991.

54. M. Bettayeb and U. AL-Saggaf, "Minimum Sensitivity IIR Filter Design Using Principal Component Approach," IEE Proc. G. (Corresp.), vol. 140, No. 4, pp. 312, Aug. 1993.

55. U. M. Al-Saggaf and G. F. Franklin, "Model Reduction via Balanced Realizations, an Extension and Frequency Weighting Techniques," IEEE Trans. on Autom. Cont., vol. AC-33, No. 7, pp. 687-692, July 1988.

56. S. Y. Hwang, "Minimum Uncorrelated Unit Noise in State Space Digital Filtering," IEEE Trans. on Acoust., Speech, and Sig. Processing, vol. ASSP-25, pp. 273-281, 1977.

57. P. Agathoklis and V. Sreeram, "Truncation Criteria for Model Reduction Using Balanced Realization ," Electr. Lett., vol. 24, No. 14, pp. 837-838, July 1988.

58. V. M. Adamjan, D. Z. Arov, and M. G. Krein, "Analytic Properties of Schmidt Parts for a Hankel Operator and the Generalized Schur-Takari Problem," Math. USSR Sbornik, vol. 15, pp. 31-73, 1971.